# DOCUMENT RESUME

ED 235 770 IR 010 825

| | |
|---|---|
| AUTHOR | Rumelhart, David E.; Norman, Donald A. |
| TITLE | Representation in Memory. |
| INSTITUTION | California Univ., San Diego. Center for Human Information Processing. |
| SPONS AGENCY | Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office. |
| REPORT NO | ONR-8302; UCSD-CHIP-116 |
| PUB DATE | Jun 83 |
| CONTRACT | N00014-79-C-0323 |
| NOTE | 126p.; This research also supported by a grant from the System Development Foundation. |
| PUB TYPE | Information Analyses (070) |
| EDRS PRICE | MF01/PC06 Plus Postage. |
| DESCRIPTORS | *Artificial Intelligence; Classification; Generalization; Literature Reviews; Mathematical Logic; *Memory; *Models; *Psychological Studies; *Schemata (Cognition); Semantics; Simulation |
| IDENTIFIERS | *Knowledge; Semantic Features |

ABSTRACT

This paper reviews work on the representation of knowledge from within psychology and artificial intelligence. The work covers the nature of representation, the distinction between the represented world and the representing world, and significant issues concerned with propositional, analogical, and superpositional representations. Specific topics related to propositionally-based systems are semantic features or attributes, symbolic logic and the predicate calculus, semantic networks and their properties, schemata and frames, and the relationship of these representations to classical associations. A section on analogical representations discusses the work of Shepard, Kosslyn, and Funt; mental models and mental simulations; and propositional and analogical representation. Procedural representation topics addressed include the human information processing system, procedural representation, and demons and production systems. Additional topics include local and superpositional memory systems, associative memories, cognition and categorization, and generalization. Major controversies within psychology--such as distinctions between declarative and procedural representation, propositional and analogical representation and the nature of visual images--are analyzed and found not to reflect fundamental disagreements. A five-page reference list is included. (Author/LMM).

*************************************************************************
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
*************************************************************************

ED235770

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

CHIP 116
June 1983

# REPRESENTATION IN MEMORY

David E. Rumelhart   Donald A. Norman

Department of Psychology

and

Institute for Cognitive Science

University of California, San Diego

La Jolla, California 92093

Copyright © 1983 David E. Rumelhart and Donald A. Norman.

Approved for public release, distribution unlimited.

Cognitive Science Laboratory

Center for Human Information Processing

University of California, San Diego

La Jolla, California 92093

IR010825

| | | |
|---|---|---|
| **REPORT DOCUMENTATION PAGE** | | **READ INSTRUCTIONS BEFORE COMPLETING FORM** |
| 1. REPORT NUMBER<br>ONR 8302 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>Representation in Memory | | 5. TYPE OF REPORT & PERIOD COVERED<br>Technical Report |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>David E. Rumelhart and Donald A. Norman | | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-79-C-0323 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Center for Human Information Processing<br>University of California, San Diego<br>La Jolla, California 92093 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>NR 667-437 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS | | 12. REPORT DATE<br>June 1983 |
| | | 13. NUMBER OF PAGES<br>117 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>Unclassified |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This research was also supported by a grant from the System Development Foundation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Representation
Knowledge
Categories
Schema
Semantic Network
Propositions
Images
Analog
Procedural
Declarative

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

OVER

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

## ABSTRACT

This paper provides a review of work on the representation of knowledge from within psychology and artificial intelligence. The work covers the nature of repesentation, the distinction between the *represented world* and the *representing world*, and significant issues concerned with propositional, analogical, and superpositional representations. Major controversies within psychology -- such as distinctions between declarative and procedural representation, propositional and analogical representation, and the nature of visual images -- are analyzed and found not to reflect fundamental disagreements. The paper is a draft of a Chapter to appear in the revision of Steven's *Handbook of Experimental Psychology*.)

# REPRESENTATION IN MEMORY

David E. Rumelhart
Donald A. Norman

University of California, San Diego

## TABLE OF CONTENTS

REPRESENTATION IN MEMORY ........ 1
- *Representations: What Are They?* ........ 1
- *Representation as Mappings* ........ 3
  - *Representation IN versus representation OF the mind.* ........ 4
- *Overview of Representational Systems* ........ 6
- *Representational Systems Include Both Representation and Process* ........ 8

PROPOSITIONALLY BASED REPRESENTATIONAL SYSTEMS ........ 10
- *Semantic Features or Attributes* ........ 10
  - *Similarity and featural representations.* ........ 13
  - *Similarity and metaphor.* ........ 15
- *Symbolic Logic and the Predicate Calculus* ........ 16
  - *The use of the predicate calculus in psychology.* ........ 18
- *Semantic Networks and Their Properties* ........ 21
  - *Inheritance properties and default values.* ........ 21
  - *The representation of n-ary relations in semantic networks.* ........ 25
  - *Types and tokens.* ........ 25
  - *Spreading activation in semantic networks.* ........ 28
  - *ACT and the fan effect.* ........ 30
  - *Schank's conceptual dependency.* ........ 32
  - *KL-ONE.* ........ 35
- *Schemata and Frames* ........ 38
  - *Summary of the major features of schemata.* ........ 41
  - *Scripts, plans and MOPS.* ........ 45
  - *KRL: A knowledge representation language.* ........ 50
- *The Relationship of These Representations to Classical Associations* ........ 54

ANALOGICAL REPRESENTATIONS ........ 55
- *Shepard* ........ 55
- *Kosslyn* ........ 58
- *Funt* ........ 68
- *Mental Models and Mental Simulations* ........ 70
  - *Mental simulation.* ........ 71
  - *The essential features of mental models.* ........ 71
- *Propositional and Analogical Representation* ........ 72
  - *Analogical does not mean continuous.* ........ 76

PROCEDURAL REPRESENTATIONS ........ 78
- *The Human Information Processing System* ........ 78
- *The Representation of Procedures* ........ 79
- *Procedural Representation* ........ 81

*Psychological implications.* .................... 83
*Actor and object based systems.* .................... 84
*Demons and Production Systems* .................... 85
*Demons.* .................... 85
*Production systems.* .................... 86
*Expert systems.* .................... 91
*Problems with production systems as models of human processing.* .................... 92
**SUPERPOSITIONAL MEMORIES** .................... 93
*Local and Superpositional Memory Systems* .................... 93
*Associative Memories* .................... 94
**GENERAL ISSUES IN THE STUDY OF REPRESENTATION** .................... 99
*Cognition and Categorization* .................... 99
*Cognitive economy and perceived world structure.* .................... 99
*Shareability constraints.* .................... 100
*Basic level categorization.* .................... 101
*How are categories defined?* .................... 102
*Prototypes.* .................... 103
*Generalization* .................... 104
*Generalization through the formation of generalized schemata.* .................... 105
*Generalization without specific generalized concepts.* .................... 105
*Superpositional models of generalization.* .................... 106
**CONCLUSION** .................... 107
*Three Major Controversies* .................... 107
*The propositional -- analogical controversy.* .................... 107
*The continuous -- discrete controversy.* .................... 108
*The declarative -- procedural controversy.* .................... 108
*Data Structure and Process* .................... 109
*Multiple Representations* .................... 109
*Virtual knowledge.* .................... 110

This is a draft of a chapter for the revision of the Steven's "Handbook of Experimental Psychology": R. C. Atkinson, R. J. Herrnstein, G. Lindzey, and R. D. Luce (Eds.), *Handbook of Experimental Psychology*. Wiley: in preparation. Comments are welcomed.

Preparation of this chapter was partially supported by Contract N00014-79-C-0323, NR 667-437 with the Personnel and Training Research Programs of the Office of Naval Research and by a grant from the System Development Foundation. Requests for reprints should be sent to the authors at the Institute for Cognitive Science C-015; University of California, San Diego; La Jolla, California, 92093, USA.

# REPRESENTATION IN MEMORY

Problems of representation are central issues in the study of memory and of cognition as a whole. Questions of how knowledge is stored and used are involved in nearly all aspects of cognition. In spite of its centrality (perhaps because of it) issues surrounding the nature of representation have become some of the most controversial aspects of the study of cognition. At the same time, representation has become one its most muddled concepts. For most Cognitive Scientists, it is impossible even to imagine a cognitive system in which a system of representation does not play a central role. But even among those for whom the concept of representation is taken to be central, there are still tremendous debates concerning the precise nature of representation:

- What is a representation anyway?
- Is it analogical or propositional?
- Is it procedural or declarative?
- Is there only one kind of representation or are there several?
- What does memorial information look like?
- Is the information stored in memory organized so that related information is stored together, or is it stored in packets or records, each independent of the remaining packets?
- Is knowledge stored as a collection of separate units or are individual memory traces intertwined over large regions of memory?

## *Representations: What Are They?*

Much of the research in Cognitive Science has been concerned with the representation of knowledge and, more particularly, the representation of meaning. The rationale goes something like this. Meaning is an important part of understanding, remembering, and cognition. If we want to make a process model of understanding or remembering or cognition, there must be something in our model corresponding to meaning. But what should meaning look like? It is natural to turn to the logicians for ideas on how to represent meaning. The major language of the logicians is the predicate calculus. Thus, most of the early ideas as to how we should represent meaning was with formulas of the predicate calculus, and so our story starts there.

Suppose that Fido were a *DOG*, and that Fido were also a *PET*. We could represent these two statements by letting *PET* and *DOG* take as arguments, a particular instance:

*A:* *DOG(Fido)*
*B:* *PET(Fido)*

Now, let $x$ be any particular instance of a *PET*. If all possible instances of *PET* were also *ANIMALS*, we would express this as

$$\forall x\ (PET(x) \rightarrow ANIMAL(x))$$

where the symbol "$\forall$" is the "universal quantifier" and is to be read, "for all": the formula then reads, for all $x$, if $x$ is a *PET*, then $x$ is an *ANIMAL*. Note that if some person $p$ owns a rock and insists that it is a pet, then the formula is false, because for $x = rock$, $PET(x)$ but not $ANIMAL(x)$. To express the fact that there is at least one $x$ that is a *PET* and not an *ANIMAL* (namely the case where $x = p$'s rock), we would say

$$\exists x\ (PET(x)\ AND\ FALSE(ANIMAL(x)))$$

where the symbol "$\exists$" is the "existential quantifier" and is to be read, "there exists:" there exists an $x$ such that $x$ is a *PET* and it is FALSE that $x$ is an *ANIMAL*.

In the early days of computer models of language understanding and semantic memory, this was a common representational format. When the predicate calculus representations were employed, the rules for operating on representations were based upon logical rules of inference. This led to the development of a number of artificial intelligence systems which employed general theorem proving programs for making inferences as a natural consequence of choosing the logician's method of representation.

To many people, the very power of logical representation was its difficulty: the predicate calculus solves problems that people find difficult, and although this is virtuous in a mathematics, it is not appropriate for a model of human thought. After all, a model of human representation should find easy what people find easy, difficult what people find difficult. However, in making this complaint, it is important not to confuse the tool with the product. The predicate calculus is a tool, with considerable explanatory and mathematical power. It is a useful means for encoding our beliefs about human representation. With it, we could model the strengths and weaknesses of human thought. Just as we can model a bouncing ball with differential equations without believing that the ball itself understands or solves these equations, we can model human processes with various formalisms without believing that the human knows about, understands, or uses those formalisms. Tools are descriptive, not explanatory. Nonetheless, in general, models of human representational processes have tended to avoid the use of the full power of the predicate calculus.

We can illustrate another kind of problem people sometimes have with these systems by relating some of the problems encountered when trying to teach some of these issues to undergraduates many years ago. The problems came up with the representational scheme used in early studies of psycholinguistics by Clark and Chase (1972), but the point is much more general than their work. Clark and Chase presented their subjects simple pictures which sometimes had a star above a plus, sometimes a plus above a star. Then, their subjects were shown printed sentences of the form

"The plus is not below the star."

and asked to respond TRUE or FALSE depending on whether the information in the sentence matched that in the picture. The details are unimportant for this illustration. The important point is that Clark and Chase assumed that subjects looked at the picture and "represented" it in the form

*(ABOVE(STAR,PLUS))*

and then represented the sentence in something like the form

*(NOT(BELOW(PLUS,STAR))).*

The judgment was thought to be made on the basis of a comparison and transformation of these two representations. In spite of the impressive fit of their model to the data, our undergraduates could not be convinced that this theory was at all reasonable.

Our students said, "We certainly wouldn't do it that way." "Why not?" we asked. "Well," they replied, "the representation is too sparse, it lacks information of 'how much' one object is above the other, and of the exact sizes and shapes of the 'plus' and the 'star.'" In short, our students felt that regardless of the impressive fit of the theory to the data, the theory was wrong because the representations did not match the richness of their personal impressions of their own representations.

What is going on? Were Clark and Chase so caught up in their narrow view of things that they missed something so obvious that any sophomore could see it? Or, were the undergraduates just too naive to understand the implications of their theories and the irrelevance of their intuitions. The real problem lies in our lack of clarity about what a representation is and about what properties a representation should have.

### *Representation as Mappings*

Let us now try to be clear about what kind of a thing a representation really is and use that to see why our students had so many problems. To begin, a representation is something that stands for something else. In other words, it is a kind of a model of the thing it represents. We have to distinguish between a *representing world* and a *represented world*. The representing world must somehow mirror some aspects of the represented world. Palmer (1978) has listed five features that must be specified for any representational system:

(1) what the represented world is;

(2) what the representing world is;

(3) what aspects of the represented world are being modeled;

(4) what aspects of the representing world are doing the modeling;

(5) what the correspondences are between the two worlds.

These features are illustrated in Figure 1. In this example the represented world consists of two stick figures -- one taller than the other. We can imagine that each has the property of having some height and the relationship *TALLERTHAN* holding between the first and second figure. We have illustrated four different possible representing worlds. In the first (I) we have the symbol *A* representing the taller figure and the symbol *B* representing the shorter. We represent the relationship among the height of the two by the formula *TALLERTHAN(A,B)*. There is no direct representation of height in this system. In the second example (II) the figures are represented by lines and height is directly represented by line length. The *TALLERTHAN* relation is implicitly represented by the physical relation *LONGERTHAN* among the line segments. In the third example (III), numbers are used to represent the figures and the magnitude of the numbers represent their heights. The *TALLERTHAN* relation is represented by the arithmetic relation of *GREATERTHAN* (>). Note that the representational format is quite arbitrary. Thus, example IV shows an alternative format for using the magnitude of numbers to represent heights, in this case, with the taller figures represented by smaller numbers; the *TALLERTHAN* relation is represented by the arithmetic relation of LESSTHAN (<). If our only goal were to represent height, then the representational systems of III and IV would be functionally equivalent. These four examples illustrate how the same characteristic in the represented world can be represented very differently in different representing worlds.

We can express these ideas more precisely. In general, a world consists of a set of objects and a set of relations among those objects. So, for example, one world, the represented world, might consist of a set of objects, *A*, and a set of relations *R*. In the formal language of relational theory this can be denoted by the two-tuple $<A,R>$. Not all aspects of the represented world are modeled in the representing world, however, so we let $A'$ and $R'$ stand for those objects and relations, respectively, that are to be represented. This subset of the to-be-represented world can be designated $<A',R'>$. In the representing world, there is a corresponding set of objects, $B'$, and a function $f$ such that for every object $a'$ in $A'$, there is an object $b'$ in $B'$, such that $f(a') = b'$. There is also a corresponding set of relations $S'$ in the representing world such that if $a'_1$ is related to $a'_2$ by relation $R'_{12}$ then $f(a'_1)$ is related to $f(a'_2)$ by relation $S'_{12}$. In other words, in a representational system, there are three relevant ordered pairs, one $<A,R>$ for the represented world, one $<A',R'>$ for those aspects of the represented world that are being modeled, and one $<B',S'>$ for what is within the representing world. There are two relevant mappings: one between objects -- $A'$ and $B'$ -- and another between relations -- $R'$ and $S'$.

*Representation IN versus representation OF the mind*. The most important point of a representation is that it allows us to reach conclusions about the thing being represented by looking only at the representing world. When considering how knowledge is represented in the human there are four kinds of things we need to keep in mind:

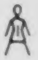
| | Represented World | Representing World | | | |
|---|---|---|---|---|---|
| | | I | II | III | IV |
| Objects: |  | A | | 15 | 7 |
| | | B | | 13 | 9 |
| Properties: | height | not directly represented | line length | numeric value | numeric value |
| Relations: | a taller than b | TALLERTHAN(A,B) | LONGERTHAN | GREATERTHAN | LESSTHAN |

Figure 1. The relationship between the represented world and the representing world showing four different ways the *representing* world might chose to model the physical-relation of *TALLERTHAN* that holds between the two figures in the *represented* world. I shows a propositional representation: *TALLERTHAN(A,B)*. II shows a representation by means of line length. III shows a representation by means of numerical value, and IV shows that the relationship can be arbitrary, as when smaller numbers in the representing world represent larger figures in the represented world.

(1) An environment in which there are objects and events;

(2) A brain which attains certain states dependent on its current state and the sensory information that impinges on it;

(3) Our phenomenal experience, which is assumed to be a function of our brain state;

(4) A model or theory of the environment, the brain states, and the experience.

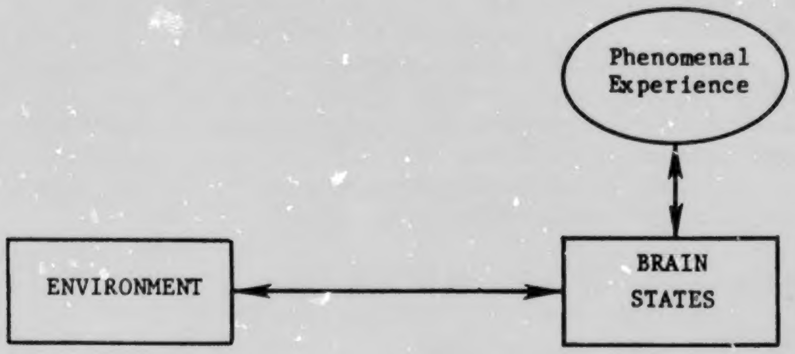
In studying representational systems, it is important to realize that there are several different pairs of representing and represented worlds, and that our *theories* of representation are in actuality representations of a representation: that is, *representations* of the mental activity that in turn is a representation of the environment. Thus, as shown in Figure 2, within the brain there exist brain states that are the representation of the environment. The environment is the represented world, the brain states are the representing world. Our theories of representation are in actuality representations of the brain states, not representations of the world. Therefore, theories of representation have the brain states as the represented world and the theoretical structures as the representing world. Finally, our phenomenal experience reflects the brain states, and so can be considered a representing world with the brain states as their represented world. When people think of representation, they often think of the relationship between phenomenal experiences and the environment, but in fact, this relationship is a secondary one, with brain states as an intermediary, although this is seldom stated explicitly in psychological theories of representation.

Presumably, our students had access to their phenomenal experience, and when they compared it with the world represented by Clark and Clark, they found their experiences richer and more complete. However, Clark and Clark only claimed to represent $A'$ and $R'$, small, limited subsets of $A$ and $R$, not the full environment. Moreover, our students were comparing their phenomenal world with a limited representing world; there is no wonder that they were unhappy. Consider the sense in which our phenomenal experience "represents" the external world. There are objects in the world and there are objects of experience. The objects of our experience are not the same as the objects of the world, but they would seem to reflect much of the structure of the world. In this way, it probably does make sense to speak of our experiential "representation" of the world.

### *Overview of Representational Systems*

The representational systems most popular today fall into four basic families. These are:

(1) The propositionally based systems in which knowledge is assumed to be represented as a set of discrete symbols or propositions, so that concepts in the world are represented by formal statements.


Phenomenal Experience
ENVIRONMENT
BRAIN STATES

REPRESENTED WORLD ⟶ REPRESENTING WORLD

The following relationships hold:

environment ⟶ brain states

brain states ⟶ Phenomenal experience

brain states ⟶ Theories of representation

The following relationship does NOT hold:

environment ⟶ Phenomenal experience

Figure 2. The relationships among the represented world, the brain, and the environment.

(2) Analogical representational systems in which the correspondence between the represented world and the representing world is as direct as possible, traditionally using continuous variables to represent concepts that are continuous in the real world. Examples are the use of electrical voltages in an analog computer to represent fluid flow or shaft rotation, or maps that are analogical representations of some geographical features of the world, or pictures in which three-dimensional space is represented by marks on a two-dimensional medium.

(3) Procedural representational systems in which knowledge is assumed to be represented in terms of an active process or procedure. Moreover, the representation is in a form directly interpretable by an action system. Consider how to pronounce the word "serendipitous." The movement made by the vocal apparatus is clearly procedural in that it is tied up in the actual performance of the skill and is not available apart from the ability to do the task, even though one normally does have conscious control and accessibility to many of the components of the task. Thus, to describe the tongue movements made in pronouncing the word, one actually has to perform the task -- that is, to say the word "serendipitous" -- and then describe aloud the actions performed.

(4) Distributed knowledge representational systems, in which knowledge in memory is not represented at any discrete place in memory, but instead is distributed over a large set of representing units -- each unit representing a piece of a large amount of knowledge.

Most actual representational systems are hybrids that fall into more than one of these four categories. Nevertheless, these categories form a useful framework within which to describe the various systems that have been proposed.

### *Representational Systems Include Both Representation and Process*

We have introduced several categories of representational systems. There is, however, one more important aspect of a representation system that must be considered: the processes that operate upon the representations. Consider the four different representational formats illustrated in Figure 1. The point of this figure was to demonstrate some of the properties of the four formats. But note that the representations within the representing world did not carry their meaning without the assistance of some process that good make use of and interpret the representational structures. Thus, if height is to be represented by line length, there must exist some process capable of comparing line lengths. If height is to be represented by numbers, then there must be some processes that can operate upon those numbers according to the appropriate rules of mathematics and the rules established by the choice of representation (e.g., whether it is type III or IV in Figure 1). Similarly, the representational system established by the use of formulas from the predicate calculus requires interpretation and evaluation. In all these cases, the processes that evaluate and interpret the representations are as important a the representations themselves.

In general, a ***Representational System (RS)*** **involves a relational double:**

$$RS = <R, P>,$$

**where *RS* is the entire system, *R* the representing world (which itself requires the several ordered pairs discussed earlier), and where *P* is the set of processes that operate upon and interpret *R*. In general, there are many forms of processes. Moreover, there is a tradeoff possible between *R* and *P*, so that information that some systems chose to include within *R* can be included within *P* by others. In some systems, the distinction between the representation (*R*) and the processes that operate upon them (*P*) is clear and distinct; in others, the *R* and *P* are so tightly intertwined that clear distinctions are impossible. In all cases, however, it is necessary always to recognize that a representational *system* is incomplete unless both the representation and the processes that operate upon them have been explicitly considered.** [1]

---

[1] **In general, the *R* part of *RS* is called the *declarative* part of the system and the *P* part is called the *procedural* part. We return to this distinction later.**

## PROPOSITIONALLY BASED REPRESENTATIONAL SYSTEMS

Most of the representational systems that have been developed and evaluated to date fall into the category of propositional representations. These representational systems all share the characteristic that knowledge is represented as a collection of symbols. According to some views these symbols are structured into trees or networks. According to other views, knowledge merely consists of lists of such symbols. According to still other views, knowledge is thought of as highly structured configurations of such symbols with associated procedures for interpreting the symbols.

In philosophy, a *proposition* is a statement that has a truth value, determined by conditions in the world. A *predicate* is a general statement; propositions are predicates with particular values substituted for the general variables of a predicate. Thus, *DOG(x)* is a predicate and is often interpreted as the set of dogs. *DOG(Sam)* is a proposition asserting that *Sam is a dog;* it is either true or false depending on the nature of *Sam*. The technical aspects of propositions and predicates have been relaxed considerably in the development of theories of representation in psychology and in artificial intelligence, most especially the requirements that a proposition have a truth value. In this section we illustrate the use of propositional representation as it has been used in psychology, proceeding from the simplest to the most complex of propositional systems. In each case, we describe the basic issues addressed by the proponents of these systems.

### *Semantic Features or Attributes*

Perhaps the simplest of the propositional representation systems is the assumption that concepts are properly represented as a set of semantic features or attributes. This means of representation is a very natural application of the language of set theory to the problem of characterizing the nature of concepts. Variations on this view have been very popular in the study of semantic memory and as assumptions describing the representation of knowledge. According to these views, concepts are represented by a weighted set of features. Thus, concepts can stand in the familiar set relationships: two concepts can be disjoint (have no attributes in common); overlap (have some but not all attributes in common); be nested (all of the attributes of one concept are included in another); or be identical (be specified by exactly the same set of features). The features can have weights associated with them that represent various saliency and importance characteristics for the concepts in question.

Rather than review all of the applications of these ideas here, we choose to describe two well developed variations on this general theme: the "feature comparison" model proposed by Smith, Shoben and Rips (1974) and the "feature matching" model of Tversky (1977; Tversky & Gati, 1978). The proposals of Smith *et. al.* were made in the context of a series of studies that began with Collins and Quillian (1969) and Meyer (1970) on simple "semantic verification" tasks. The general procedure followed in these studies was to present a statement that asked whether a member of one semantic category could also be a member of another. Thus, typical sentences would be: *A robin is a bird, A vegetable is an artichoke,* or perhaps, *A rock is a furniture.*

Subjects were asked to respond "TRUE" or "FALSE" to the sentences as quickly as possible. The basic representational assumption was that the words representing the two categories to be considered could be represented by a set of semantic features that vary in their relationship to the formal definition of the category. In particular, features could be divided into those that were "defining" (they must hold if an item is a member of the category) and those that were "characteristic" (they usually apply, but are not necessary for the definition). Thus, *has feathers* is a *definitional feature* for the concept bird, whereas *can fly* is a *characteristic feature;* birds characteristically can fly, but flying is not essential to a thing being a bird. In addition, the concept bird might have features specifying that is has a particular size, shape, etc., things that might be true of only the most typical instances of birds. Figure 3 (from Smith & Medin, 1981) shows an illustrative set of features and weights for the concepts of *robin, chicken, bird,* and *animal.*

In formulating their proposal, Smith *et. al.* had a number of empirical results in mind. Collins and Quillian (1969) found that subjects took less time to verify statements of the form *A canary is yellow,* than statements of the form *A canary has feathers,* which in turn took less than the time to verify *A canary eats food.* From this they deduced that the information is stored hierarchically; properties specific to canaries are stored with the concept canary, properties specific to birds in general are stored with birds, and properties specific to animals are stored with animals. Thus, the further up the hierarchy one has to search to find the relevant information, the longer it takes subjects to answer the question. Smith *et. al.* found that the time to verify a statement does not always conform with the predictions from a hierarchical model. Thus, it might take longer to confirm that *A cat is a mammal* than to confirm that *A cat is an animal.* More interestingly, it was found that it is faster to verify that *A robin is a bird* than to verify that *A chicken is a bird* or that *A penguin is a bird* (Rips, Shoben & Smith, 1973). In general, the more *typical* an instance is of a category, the more quickly it can be verified that it, in fact, belongs to that category.

Smith, Shoben, and Rips (1974) proposed that category membership is not a prestored characteristic but rather was computed from the comparison of a set of features. They proposed that the process of verifying a category membership statement consisted of two stages. First, a very quick comparison of all features (characteristic and defining) was performed. If this comparison was sufficiently good, the question was answered in the affirmative. If the comparison was sufficiently poor, the question was answered in the negative. If the comparison led to an intermediate result, a slower comparison process applied to the defining features was initiated. This model accounts for the basic experimental results: true statements involving highly typical items (e.g., *A robin is a bird)* are affirmed very quickly; false statements involving very distinct items (e.g., *A door is a bird)* are rejected very quickly; statements involving less typical examples of a category (e.g., *A penguin is a bird)* are affirmed relatively slowly; and statements involving things similar to, but not members of, the category (e.g., *A bat is a bird)* are rejected relatively slowly.

A number of different kinds of verification proposals have been made, all somewhat different from one another, but all consistent with the spirit of this general approach. Thus, McCloskey and Glucksberg (1979) employ similar assumptions about

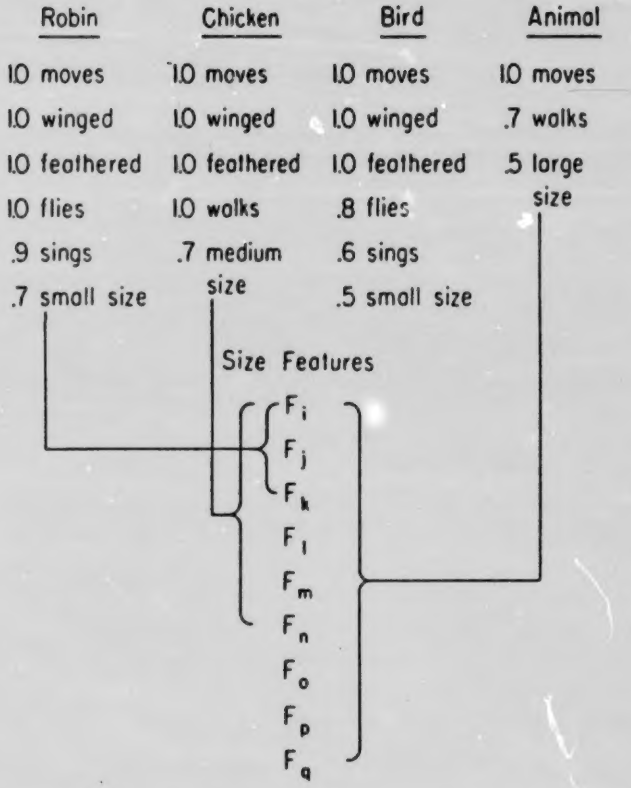

Robin
1.0 moves
1.0 winged
1.0 feathered
1.0 flies
.9 sings
.7 small size
Chicken
1.0 moves
1.0 winged
1.0 feathered
1.0 walks
.7 medium size
Bird
1.0 moves
1.0 winged
1.0 feathered
.8 flies
.6 sings
.5 small size
Animal
1.0 moves
.7 walks
.5 large size
Size Features
$F_i$
$F_j$
$F_k$
$F_l$
$F_m$
$F_n$
$F_o$
$F_p$
$F_q$

Figure 3. An illustrative set of features and weights for the concepts of *robin, chicken, bird*, and *animal* (from Smith & Medin, 1981).

representation of concepts, but only require a single stage comparison process. The newer models, of course, usually account for the data better than do the earlier models. The important point, however, is that all of these models assume that conceptual knowledge is represented by a set of features and that these features include necessary and sufficient attributes of the concept being represented as well as attributes that are only characteristic of the concept being represented. Because the category contains features that are typical of its instances, but that are not necessarily shared by its instances, these models are referred to as *prototype* theories of representation.

*Similarity and featural representations*. Judgements of the similarity of two concepts pose a particularly interesting problem. The most obvious way to approach the problem is to state that two concepts are similar inasmuch as their underlying features are similar, or overlap. If each concept is represented by a set of $N$ features, then one can think of the features as representing an $N$ -dimensional space, with each of the concepts being a point in the space, with location specified by the weights or values of each concept in the feature definition. In models of this type, similarity is often assumed to be a monotonically decreasing function of the distance between points in the multidimensional space. Any geometric representation of this form must satisfy two major conditions: symmetry and the triangular inequality. The symmetry condition states that because similarity is a function of the distance between points, the similarity of $A$ to $B$ must be the same as the similarity of $B$ to $A$. The second condition, the triangle inequality condition, states that for any three points, the distance between any two must be less than or equal to the sum of the distances between the other two. Because similarity is inversely related to the distance between points, the triangular inequality translates into the condition that the similarity of two concepts $A$ and $C$ must be greater than or equal to the sum of the similarity of $A$ to $B$ and of $B$ to $C$. Both these basic properties may be violated (Tversky, 1977; Tversky & Gati, 1978).

Tversky points out that in certain cases similarity appears to be an asymmetric relation. For example, people generally judge the similarity of North Korea to Mainland China to be greater than the similarity of China to North Korea, thus violating the symmetric property. The triangular inequality can also be violated. Thus, although Jamaica is very similar to Cuba (due to its geographical characteristics) and Cuba is similar to Russia (politically), Jamaica is not at all similar to Russia.

Tversky suggests that these violations can be readily accounted for by means of a simple model defined on a semantic feature representation. Tversky's major representational assumptions are essentially identical to those of Smith, Shoben, and Rips (1974). Figure 4 shows the relationships between the representations of two overlapping concepts a and b. Note, there are seven sets of features distinguished in this relationship. These are:

(1) The features of concept *a:* the set *A;*

(2) The features of concept *b:* the set *B;*

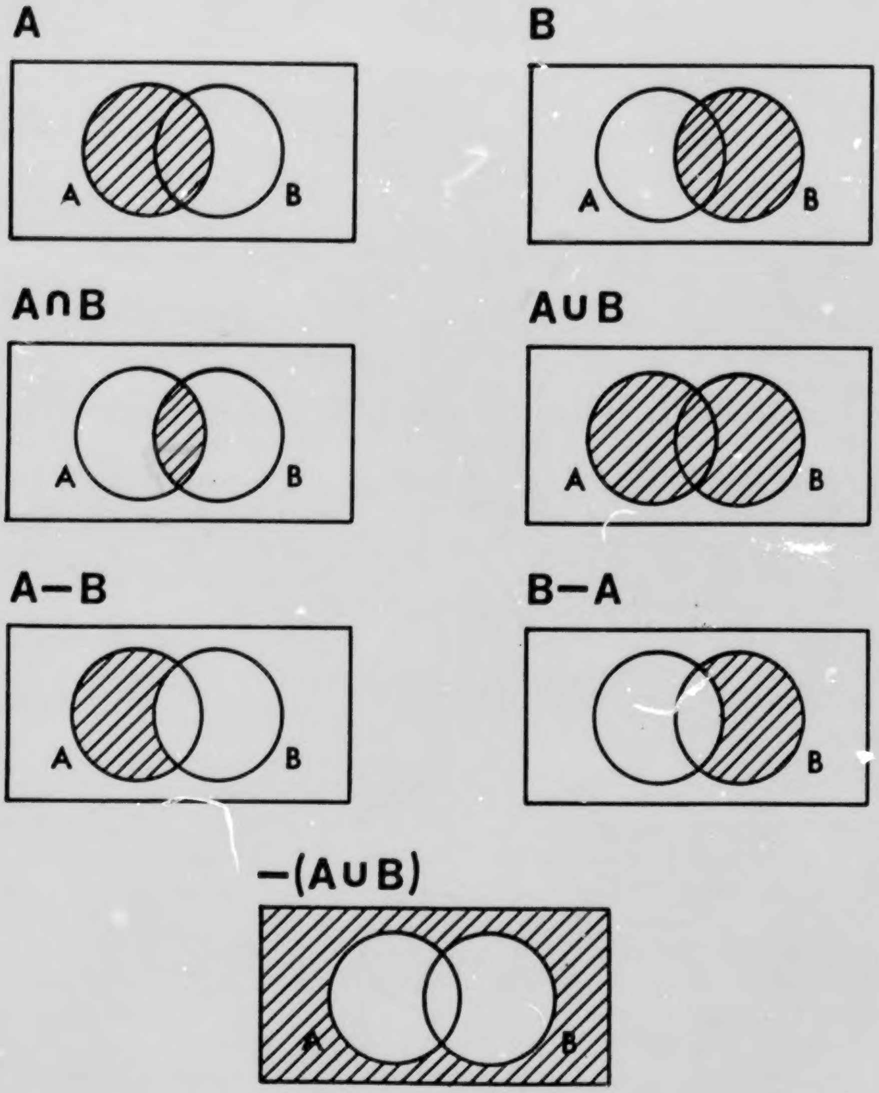


Figure 4. The seven different relationships that can apply among members of two overlapping sets *(A* and *B).* They may be members of one set *(A)* or of the other *(B).* They may be in common between the two sets *(A∩B).* They may be in either *A* or *B* *(A∪B).* They may be in *A*, but not in *B (A–B)* or they may be in *B*, but not in *A (B–A)*, and finally, they might be in neither *A* nor *B (–(A∪B)* or *(–A)∩(–B)).*

(3) The features common between $a$ and $b$: the set $A \cap B$;

(4) The total set of features either in $A$ or in $B$: the set $A \cup B$;

(5) The features that are in $a$ but not in $b$: the set $A-B$;

(6) The features that are in $b$ but not in $a$: the set $B-A$; *and*

(7) The features that are neither in $A$ nor in $B$: the set $(-A) \cap (-B)$;

Tversky proposes that that the similarity of $a$ to $b$, $S(a,b)$, be given by the equation

$$S(a,b) = f(A \cap B) - \alpha f(A-B) - \beta f(B-A)$$

where $f(X)$ is a measure of the salience of the features in set $X$ and $\alpha$ and $\beta$ are constants. Tversky's account of similarity suggests that different aspects of the representation are treated differently, depending upon the question being asked. Thus, if $\alpha > \beta$, then $a$ is more similar to $b$ than is $b$ to $a$: $S(a,b) > S(b,a)$ (as in the China-Korea example). If the weights associated with the different dimensions change during the answering of the question to reflect the different dimensions being considered, then such properties as the triangular inequality can be violated (as in the Jamaica, Cuba, Russia example).

*Similarity and metaphor.* Ortony (1980) has applied Tversky's model to the similarity of metaphorical statements such as:

*Lectures are like sleeping pills.*

*Sleeping pills are like lectures.*

*Lectures are like sermons.*

Like Tversky, Ortony noted an extreme asymmetry in the meaning of these statements. The first seems to be an altogether reasonable (albeit metaphorical) assertion whereas the second seems to be nearly nonsensical. On the other hand, the third seems to be a straightforward statement of literal similarity. Following Tversky, Ortony suggests that the meaning of the concepts "lectures" and of "sleeping pills" are represented by sets of features, each with an importance or salience value. The meaning of these statements can be determined by matching the features of the predicate term with those of the subject term. In a normal, declarative sentence, highly salient features of the predicate term are also highly salient features of the subject term, as in the third example. A sentence is metaphorical or a simile if highly salient predicate features are relatively low salient subject features. Finally, sentences of this form are nonsensical if the subject and predicate either have no features in common or if only features that are low in salience on the predicate term are held in common.

In spite of their relative simplicity, semantic feature models offer remarkably good accounts of a rather wide body of data. (A good review of these issues is presented in Smith and Medin, 1981.) Such theories do, however, have their limitations. In particular, almost all of the work has been with simple nominal concepts. It is much less clear how these models would be applied in the case of predicate concepts. Similarly, it is not clear how such models would represent simple facts (e.g., *typewriters are used for typing)* or simple events (e.g., *John went to the store).* The semantic feature model does not handle distinctions between the statements that a robin is a bird, a sparrow is a bird, but that a sparrow is not a robin: if category membership were determined solely by defining characteristics, one might very well determine that a sparrow was a robin, or perhaps that a bird was a robin. In similar fashion, these models cannot account for problems of quantification, as represented in the contrast in meaning between the sentences *Everyone kissed someone* and *Someone was kissed by everyone*. In fairness to semantic feature models, they were not intended to solve all of the problems of representation, but rather primarily those of similarity and of definition. In this, they do well. In interpreting the role of this class of models, it is useful to note Tversky's comments on the matter (which are also relevant to the dilemma faced by our poor undergraduates who felt that these representations were lacking in substance):

> Our total data base concerning a particular object (e.g., a person, a country, or a piece of furniture) is generally rich in content and complex in form. It includes appearance, function, relation to other objects, and any other property of the object that can be deduced from our general knowledge of the world. When faced with a particular task (e.g., identification or similarity assessment) we extract and compile from our data base a limited list of relevant features on the basis of which we perform the required task. Thus, the representation of an object as a collection of features is viewed as a product of a prior process of extraction and compilation. (Tversky, 1977, p. 329).

In other words, Tversky is actually making no committment to a feature set as *the* mechanism for the representation of knowledge in general, but rather merely contends that the feature representation is produced for the purpose of carrying out particular tasks. Tversky is not pretending to offer a proposal for the representation of knowledge in general. Rather, he provides a nice account of how a feature based representation could solve the knotty problem of similarity.

### *Symbolic Logic and the Predicate Calculus*

The semantic feature representations were directed at the representations of word meanings. To represent knowledge in general we must be able to represent the meaning of arbitrary statements as well as the meaning of single words. When psychologists, linguists, and computer scientists began to concern themselves with this more general task, it was natural to look to the formalisms already developed for this purpose by mathematicians and logicians -- namely, symbolic logic. In particular, a number of workers have been drawn to the predicate calculus (developed first by Frege, 1892) as an appropriate representational format for meaning in general. On this view the representational system consists of five kinds of entities:

*Constants* (designated $a, b, c, \ldots$), expressions that stand for individual objects. Examples: proper names, such as Fido or John.

*Variables* (designated $x, y, z, \ldots$), expressions that stand for some one of a set of constants, as in, for some $x$, such that $x$ is a person.

*Predicates* (designated $P(x,y, \ldots)$), expressions that stand for particular properties or relations among objects. $P$ stands for some particular property, and $x$ and $y$ for variables. Example: *ATE(x,y):* some object $x$ ate some object $y$.

*Propositions* (designated $P(a, b, \ldots)$). Propositions are predicates in which particular constants have been substituted for the variables. When this occurs, we say that the predicate has been "instantiated." Propositions have truth values: the statement encoded by the proposition is either "true" or "false." Example: The predicate *ATE(x,y)*, which, when instantiated by Elaine and sandwich forms the proposition *ATE(Elaine, sandwich)* -- Elaine ate a sandwich -- which is either true or false.

*Functions* (designated $f(x,y, \ldots)$), expressions containing variables, that, when instantiated, form complex constants. Example: *TEACH(agent, recipient, locative, time)* which, when instantiated by appropriate constants might become *TEACH(Don, graduate students, conference room, Monday noon)*, representing the sentence "Don teaches the graduate students in the conference room, Monday at noon."

*Quantifiers*, including the existential quantifier, $\exists$ (there exists an $x$) and the universal quantifier, $\forall$ (for all $x$).

*Logical connectives* consisting of negation (~), conjunction ($\cap$), disjunction ($\cup$), and implication ($\rightarrow$). These connectives can combine predicates and propositions to produce more complex predicates and propositional expressions.

Consider how we might represent a few simple statements in the predicate calculus. First, consider the statement *John loves Mary*. In the predicate calculus formalism this becomes

*LOVES(John,Mary)*

Now consider the representation of *Someone loves Mary*. This would be represented as $\exists x(LOVES(x,Mary))$. In words, this formula says there exists an $x$ such that $x$ loves Mary. The $x$ in the quantifier is said to *be bound to* the $x$ in the predicate. Consider the statement *Everyone loves themselves*. This would be represented $\forall x(LOVES(x,x))$ In words, for all $x$, $x$ loves $x$. Finally consider the statements *Everyone loves someone* and *Someone is loved by everyone*. In the predicate calculus formalism, these two would be represented by

$$\forall(x)\exists y(LOVES(x,y)) \text{ and } \exists y\ \forall(x)\ (LOVES(x,y)).$$

Note that, in the first form, a different $y$ can be chosen for each $x$. The existential quantifier is said to be within the *scope* of the universal quantifier. In the other, the universal quantifier is within the scope of the existential quantifier. Thus the difference in meaning between these two sentences is a matter of scope. Finally, consider the predicate calculus translation of a sentence of the form *All men are mortal*. This is translated to be

$$\forall (x)\ (MAN(x) \rightarrow MORTAL(x)).$$

The great advantage of the predicate calculus is the large body of logical, philosophical and mathematical work that it calls upon. Many issues of representation, especially those involving quantification and logical connectives, have already been answered. The predicate calculus and versions of it have been extremely popular as a representational device in philosophical and linguistic treatments of meaning, in attempts to represent and reason with semantic information in artificial intelligence, and in psychological attempts to represent knowledge. Thus textbooks of methods in Artificial Intelligence sometimes suggest the use of the predicate calculus as a basic tool for the field (Nilsson, 1980).

*The use of the predicate calculus in psychology*. One example of the use of the formalism of the predicate calculus in psychology is given by the work of Kintsch and his colleagues. It should be noted that Kintsch explicitly disavows the general version of the predicate calculus. Kintsch (1972) argues that:

> The formalism that appears to be best suited for the task is some kind of low-order propositional calculus. I say low-order calculus because the attempt to translate language expressions into something like a fully quantified predicate calculus is surely misguided. Formal logic was developed precisely because language is so sloppy that it is insufficient for certain purposes (such as formal reasoning). To propose formal logic as a model for language only means forcing language into an intolerable straight-jacket . . . . What we need is a greatly less powerful and elegant formalism that permits the operation of lexical inference rules as well as the semantic-syntactic rules that are necessary to produce sentences, but that does not impose more order than there is. (Kintsch, 1972, p. 252)

Kintsch and his colleagues have looked at the representation of an interrelated set of sentences treating text as a "connected, partially ordered list of propositions." The predicates are concepts named by English verbs and the constants are other concepts, named either by English nouns or by other propositions. The variables of the predicates have associated labels indicating the *role* that the argument plays in the whole proposition. These role names are, by and large, drawn from the case grammar of Fillmore (1968) and consist of things such as *agent, object, recipient, instrument, source,* *goal*, etc. (see Figure 5). In Figure 5, the individual propositions are numbered and when a given proposition serves as an argument for another, the number of the embedded proposition is given. The roles, when named, are indicated prior to the argument in each proposition. Note that the same argument appears in several propositions. Kintsch argues that the interconnection of propositions through shared arguments is a necessary condition for coherence of a text.

Although the predicate calculus approach to representation has the strong advantage of providing a consistent and powerful representational structure with a well worked out inferential component, it is nevertheless not the universal choice. There seem to be several reasons why many workers in the field have chosen other alternatives. The two most important of these involve, first, issues surrounding the organization of knowledge in memory and the notion that the logical theorem proving processes so natural to the predicate calculus formalism do not seem to capture the ways people actually seem to reason. When one wishes to define processes operating on these representations other than the ones most obvious for the predicate calculus, alternative representational systems may prove more useful. Thus, many authors have chosen representational systems in which the knowledge pieces were connected to each other to form an associative network of interrelated pieces of knowledge. In this way the organization of information in memory is more perspicuously represented. Moreover, there has been a push to develop knowledge representation systems in which heuristic reasoning processes more like those we see in our subjects are easily definable.

Although the predicate calculus led the way, probably the most important work on representation for psychology has emphasized different aspects of knowledge than the formal issues of statements and quantification addressed by the calculus. Psychologists and workers in Artificial Intelligence have to a large extent explored representations that emphasized what could be thought of as the most salient psychological aspects of knowledge:

- The associative nature of knowledge;
- The notion of knowledge "units" or "packages," so that knowledge about a single concept or event is organized together in one functional unit;
- The detailed structure of knowledge about any single concept or event;
- That it is useful to consider different levels of knowledge, each level playing a different organizational role, and with higher order units adding structure to lower order ones;
- The everyday reasoning of people, in which "default" values seem to be substituted for information that is not known explicitly, in which information known for one concept is applied to other concepts, and in which inconsistent knowledge can exist.

These beliefs have guided studies of representation towards structures called semantic

Fragment of an Episode from a Short Story and the Corresponding Text Base*

| Text | Text base |
|---|---|
| This Landolfo, then, having made the sort of preliminary calculations merchants normally make, purchased a very large ship, loaded it with a mixed cargo of goods paid for out of his own pocket, and sailed with them to Cyprus. (The episode continues with a description of how this endeavor finally resulted in Landolfo's ruin.) | 1(PURCHASE,agent:L,object:SHIP)<br>2(LARGE,SHIP)<br>3(VERY,2)<br>4(AFTER,1,5)<br>5(CALCULATE,agent:L)<br>6(PRELIMINARY,5)<br>7(LIKE,5,8)<br>8(CALCULATE,agent:MERCHANT)<br>9(NORMAL,8)<br>10(LOAD,agent:L,goal:SHIP,object:CARGO)<br>11(MIXED,CARGO)<br>12(CONSIST OF,object:CARGO,source:GOODS)<br>13(PAY,agent:L,object:GOODS,instrument:MONEY)<br>14(OWN,agent:L,object:MONEY)<br>15(SAIL,agent:L,object:GOODS,goal:CYPRUS) |

*Modified from Kintsch (1976).

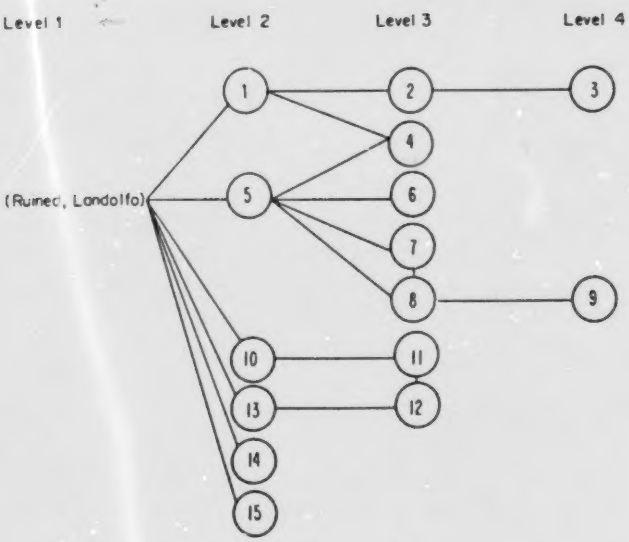


Figure 5. The text base hierarchy for the fragment of text shown at the top of the figure. Propositions are indicated only by their number; shared arguments among them are shown as connecting lines. (From Kintsch, 1978.)

networks, schemata, frames and scripts. These concepts are actually closely related to the formalisms of the predicate calculus, and in some cases are simply notational variations on the calculus. The difference in emphasis, however, is critical, for the emphasis puts the focus on functional aspects of representation, including just how a real, working system might be able to use the information. Historically, these approaches to the study of representation started with semantic networks, so let us start there as well.

### *Semantic Networks and Their Properties*

An important step in the representation of the associations within long term memory was Quillian's (1968) development of the "semantic network." The basic notion is that knowledge can be represented by a kind of directed, labelled graph structure in which the basic structural element is a set of nodes interrelated by relations. Nodes represent concepts in memory. A relation is an association among sets of nodes. Relations are labeled and directed. In this view the *meaning* of a concept (represented by a node) is given by the pattern of relationships among which it participates. It is important to note that not all nodes in a semantic memory system have names corresponding to words in natural language. Some nodes represent concepts which have no natural language equivalent, others represent instances (or tokens) of the concepts represented by other nodes. Thus, Figure 6 shows one form of network that evolved form the work of Quillian: his representation of the concept "plant" in its various meaning senses.

*Inheritance properties and default values*. One of the attractive features of the semantic network formalism is the convenience with which the property of *inheritance* is formulated. Figure 7 illustrates a common semantic network representational format for information about animals. The basic structure of a network is illustrated in the figure. Nodes (the dots and angle brackets) stand for concepts: relations (the lines with arrows) stand for the relationship that applies between the nodes. The arrows are important for specifying the direction of the relation. Any given relationship between nodes can be represented by a triple consisting of the two nodes (let them be *a* and *b* ) and the relation (let it be *R* ). In the network, the relationship is shown graphically as $a \textrm{–} R \rightarrow b$. It can also be stated in a formula, either in infix notation as *aRb* or in the more standard predicate calculus prefix notation as *R(a,b)*. We will use all three notations, for all are equivalent, but are useful at different times. Note that at any node, *a*, there may be a number of relations to other nodes, which is indeed how the network figures get constructed. [2]

---

[2] The semantic network, as drawn in Figure 7 is attractive in suggesting the kinds of inter-relations that occur among the entire set of concepts in memory and suggesting processing strategies. However, the notation becomes clumsy and unwieldy as the network structures become large and complex. Today, it is more usual to list each unit separately, putting it into what amounts to an outline form. Thus, the information in Figure 7 can be depicted in this way:

| | | | | | |
|---|---|---|---|---|---|
| animal | | | person | | |
| | eats | food | | subset | animal |
| | breathes | air | | has-as-part | legs |
| | has | mass | | has-as-part | arms |
| | has-as-part | limbs | | | |

The relation-node pairs (e.g., eats food) are called *slots* and *fillers*.

PLANT.1. Living structure which is not an animal, frequently with leaves, gettingits food from air, water, earth.
2. Apparatus used for any process in industry.
3. Put (seed, plant, etc.) in earth for growth.

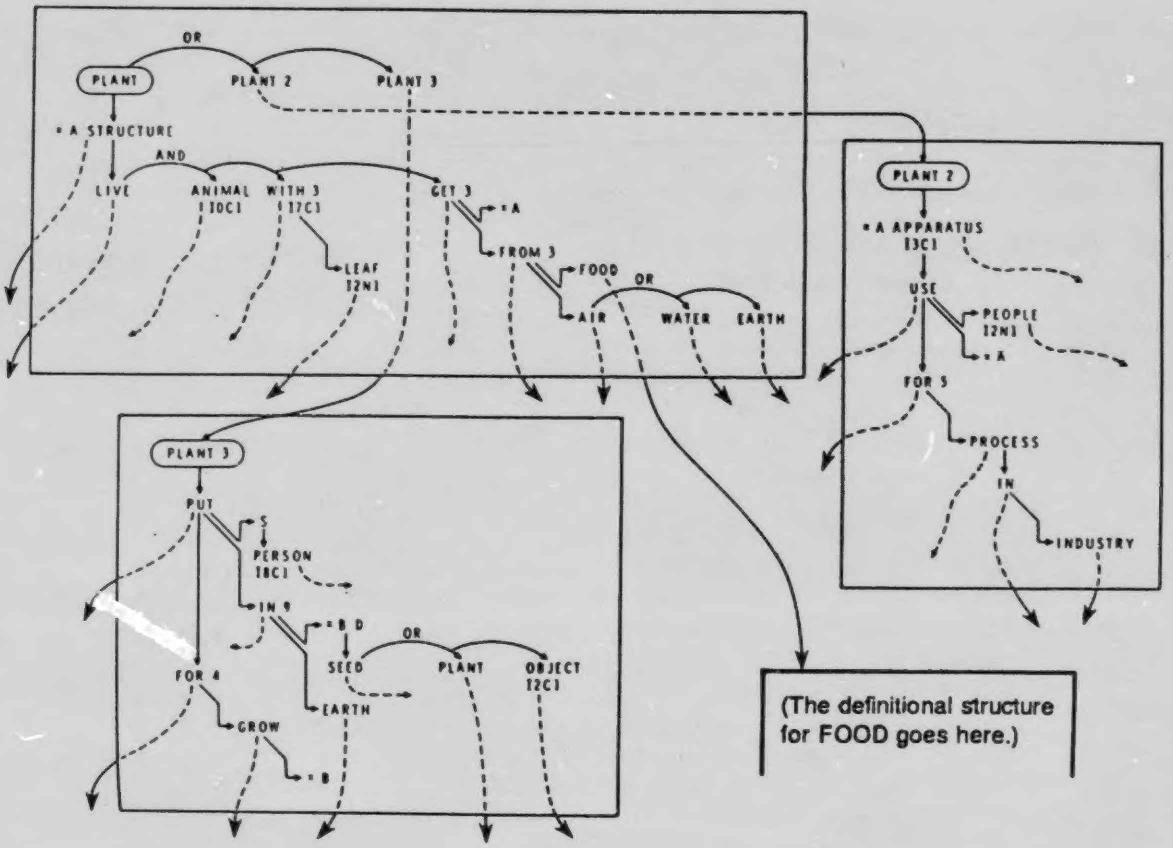

OR
PLANT
PLANT 2
PLANT 3
= A STRUCTURE
AND
LIVE
ANIMAL
WITH 3
GET 3
= A
LEAF
FROM 3
FOOD
OR
AIR
WATER
EARTH
PLANT 2
= A APPARATUS
USE
PEOPLE
= A
FOR 5
PROCESS
IN
INDUSTRY
PLANT 3
PUT
S
PERSON
IN 9
= B D
OR
SEED
PLANT
OBJECT
EARTH
FOR 4
GROW
= B
(The definitional structure for FOOD goes here.)

Figure 6. Quillian's (1968) semantic network representation for three meanings of the concept "plant."

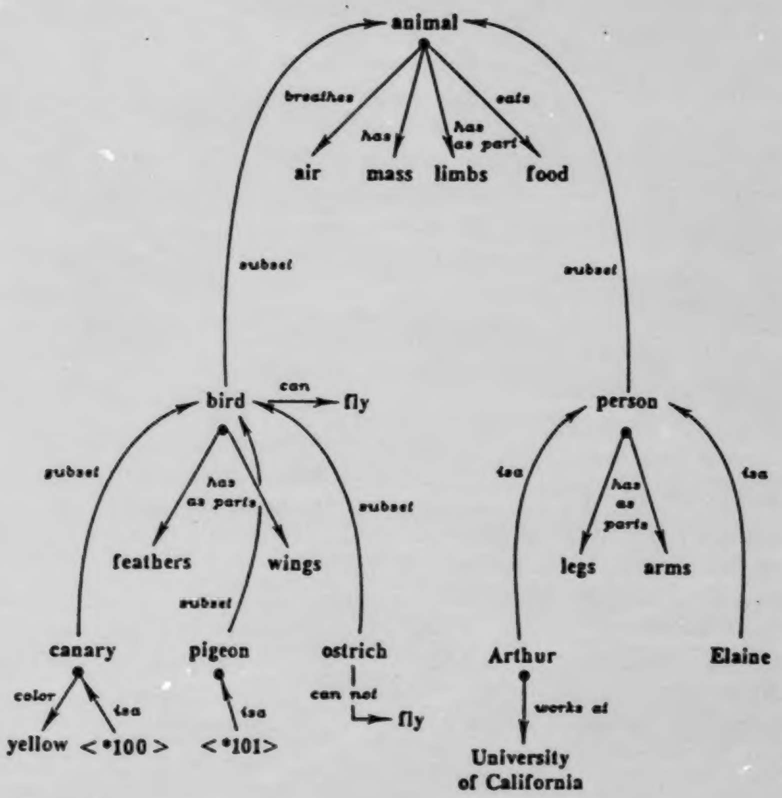

animal
breathes
has
has as part
eats
air
mass
limbs
food
subset
subset
bird
can
fly
person
subset
has as parts
subset
isa
has as parts
isa
feathers
wings
legs
arms
subset
canary
pigeon
ostrich
Arthur
Elaine
color
isa
isa
can not
fly
works at
yellow
<*100>
<*101>
University of California

Figure 7. A simple semantic network, chosen so as to illustrate the use of inheritance.

There must exist a basic set of nodes and relations, the fundamental structures that are necessary for the semantic network to work properly. An important class of relations is that of *type*, indicating that one node is an instance of the class pointed to by the relation. The two most important kinds of type relations are *isa* (where *a isa b* means that the concept represented by node a is an instance of the concept represented by node *b)* and *subset* (where *a subset b* means that the concept represented by node *a* is a subset of the concept represented by node *b)*.

Suppose we wish to represent information about animals, as shown in Figure 7. We know that animals breathe, have mass, and eat food. This information is represented by relations from the node named "animal." We know that people are animals, that Arthur and Elaine are instances of people, that birds are animals, and that canaries and pigeons and ostriches are kinds of birds. We also have seen particular birds, indicated by nodes <*100> and <*101> (indicated by angle brackets and arbitrary names). Note that the fact that Arthur eats food is derivable from the triples (Arthur isa person), (person subset animal), and (animal eats food). This derivation illustrates the property of *inheritance: instances and subsets inherit the properties of their types*. The general rule is that

*If (a type b) and (b R c), then (a R c)*

(both "isa" and "superset" are relations of class "type"). Note also that because the node for "bird" indicates that birds have feathers and fly, by inheritance, we know that these properties apply to all birds, including all of the ones in Figure 7 (canaries, pigeons, ostriches, <*100>, and <*101>). When information is applied in this way, it is called a *default value*. That is, in the absence of other knowledge, we assume (deduce) that all birds have feathers and fly. In this case, the defaults for birds is wrong: ostriches don't fly. The solution is to add to the node for ostrich that it doesn't fly (as is done in the figure). But now we have inconsistent data in the data base. In semantic networks, the issue presents no difficulty if the appropriate processing rules are followed:

1. In determining properties of concepts, look first at the node for the concept.
2. If the information is not found, go up one node along the "type" relation and apply the property of inheritance.
3. Repeat 2 until either there is success or there are no more nodes.

This processing rule will always find the lowest (most specific) level relationship that applies to a given concept and will never even notice inconsistencies of the sort illustrated in the figure. The basic principle is that if two pieces of conflicting information appear to apply to a concept, accept the one that is most specific to that concept. This basic rule turns up frequently in the application of knowledge representation to applied problems.

Semantic networks provide a convenient and powerful formalism for representing knowledge, allowing for both inferential mechanisms and processing considerations. The nice thing about the network structure is that it matches many of our intuitions for the representation of a large domain of our knowledge.[3]

*The representation of n-ary relations in semantic networks.* We have shown how the semantic network representation builds upon the node-relation-node triple *(a R b)*. Because any node can have an indefinite number of relations from it to other nodes, it is also possible to view the representation as an *n*- place predicate that applies to the concept specified by the node. In particular, if the node specifies an *n*-place predicate (a predicate with *n* arguments), then the node name can be identified with the predicate name. Each of the nodes pointed to by the relations leaving the node can be considered to be the arguments of the predicate. The relations specify the interpretation of each argument. This conceptualization makes it easy to represent complex verbs within the network, and was the scheme adopted by the LNR research group (Norman & Rumelhart, 1975). In this case, then, the basic representational unit, like that of the predicate calculus, consists of a predicate and it's associated arguments. Figure 8 illustrates the basic scheme for representing an *n*- place predicate. The central node in Figure 8*A* represents an instance or token of the predicate *P*, the labels on the relations represent the roles played by the various arguments of the predicate and the relations labeled *type* shows that this central node is a token of type *P*. Often, this structure is abbreviated as in Figure 8*B*.

*Types and tokens.* In a semantic network it is essential to distinguish between *types* and *tokens* of the concepts being represented. Figure 9*A*, illustrates the kinds of confusion that arises from failure to make the distinction. This figure is intended to represent the facts that "Cynthia threw the ball" and that "Albert threw the book." Notice that because there is only one node for "threw" we are unable to determine who threw the ball and who threw the book. Figure 9*B* correctly represents the distinction between the events of Cynthia's throwing and Albert's throwing by introducing *token* nodes, illustrated by the ovals in the figure. These token nodes are instances of the type node for "threw," allowing us to distinguish the various incidents in which the action occurs from one another.

A similar situation occurs with concepts, such as "ball." Thus, as shown in Figure 9*C*, when both Cynthia and Albert start throwing balls, we cannot tell from the representation whether or not they are throwing the same ball. We need to be able to represent that Cynthia threw a particular ball and that Albert threw some other particular ball. Basically, we use the type relation *isa*, to point from a node that represents a token instance of a concept to the node that represents its more general, type concept. (The relation "isa" can be read as "is an instance of.") Figure 9*D*

---

[3] Note, however, that semantic networks fail to capture our intuitions of the phenomenology of mental structures. In particular, their information structures do not seem to be sufficiently dense to represent the rich, perceptual and motoric component of much of our internal experiences and mental images. We return to this issue later, when we treat images.

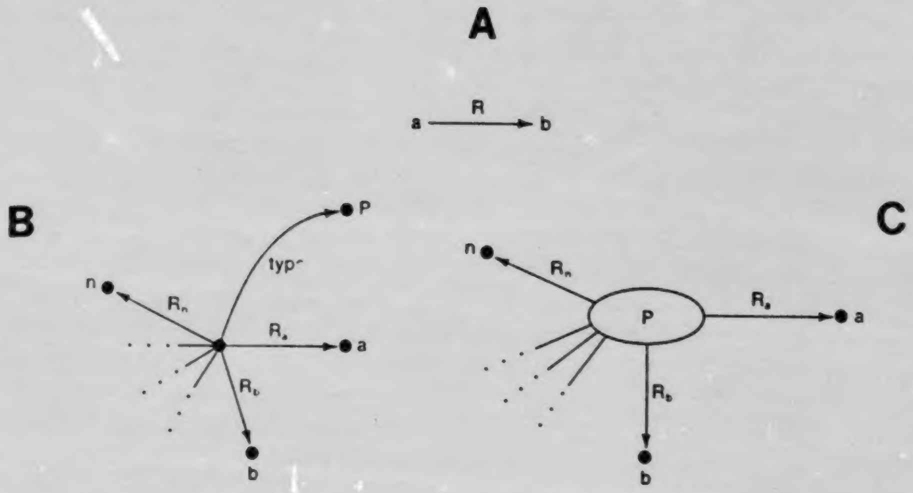


Figure 8. The basic scheme for representing an *n*- place predicate (a predicate with n arguments). The central node in A represents an instance or token of the predicate *P*, the labels on the relations represent the roles played by the various arguments in the predicate and the relation labeled *type* shows that this central node is a *token* of type *P*. An abbreviated notation is shown in *B*. When this notation is used, the connection between the node and the name of the predicate is not always shown. (From Norman and Rumelhart, 1975, p. 36.)

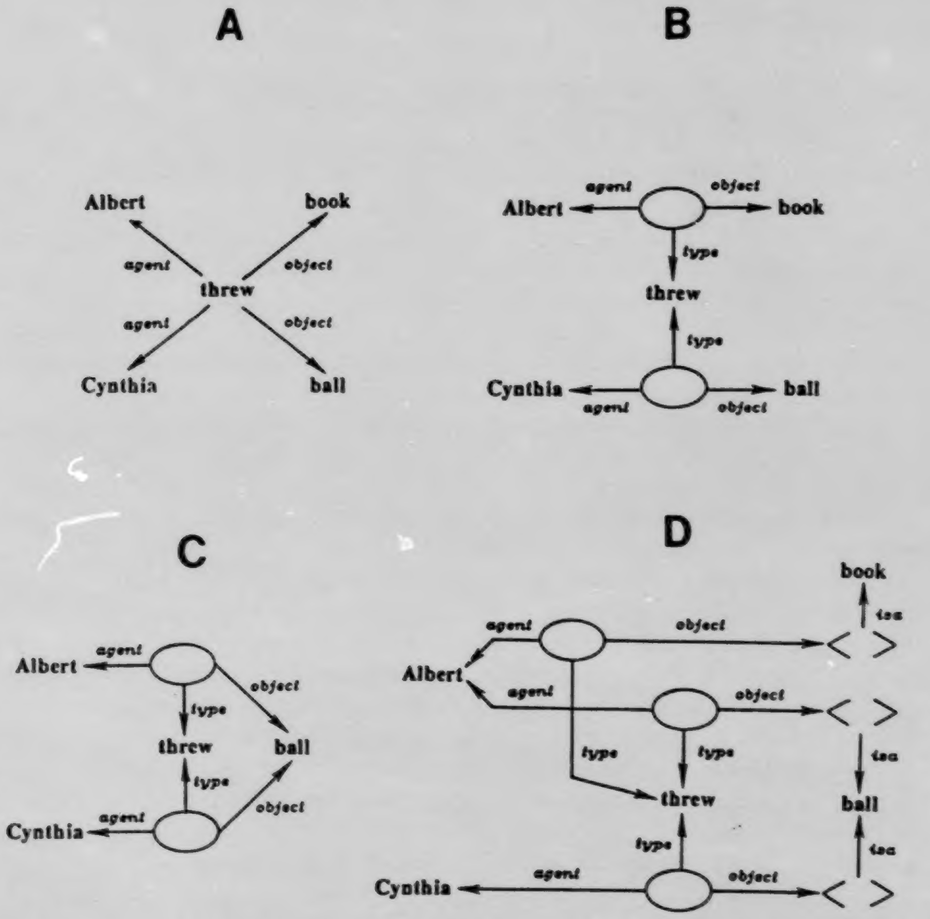

A
Albert
book
agent
object
threw
agent
object
Cynthia
ball
B
agent
object
Albert
book
type
threw
type
Cynthia
ball
agent
object
C
agent
Albert
object
type
threw
ball
type
object
agent
Cynthia
D
book
isa
agent
object
< >
Albert
agent
object
< >
type
type
isa
threw
ball
type
isa
Cynthia
agent
object
< >

Figure 9. The need for distinguishing types from tokens.

A. Who threw the ball and who threw the book?
B. Token nodes for "threw" solve the problem shown in *A*.
C. Did both Albert and Cynthia throw the same ball?
D. Token nodes for "ball" solve the problem in *C*.

illustrates how this is done, using angle brackets to represent tokens of concepts. (In most actual drawings, the "type" or "isa" relations are not shown, but the use of angle brackets and ovals indicates that the nodes are tokens and that type relations exist, but are not shown.)[4]

*Spreading activation in semantic networks.* One important processing method that has commonly been associated with semantic networks is that of "spreading activation" in which the network itself conducts activation values among its links. The first description of a spreading activation mechanisms was made by Quillian, and the ideas were most fully described and elaborated in a paper by Collins and Loftus (1975). Anderson (1976) has used it as the basis of his modeling of human memory, both for guiding psychological predictions and experimentation and also for the construction of his computer simulation.

The basic idea of spreading activation is rather simple. The semantic network is a highly interconnected structure, with relations connecting together nodes very much like highways and airline routes interlink cities of the world. Much as motor vehicles and aircraft ply the routes among cities, activation is thought to travel the routes between nodes. The concept of activation is a general one. If the model is thought of as being only a functional description, not necessarily dictating the physical system within which it is embedded, then the nodes and relations are thought of as data structures with the relations being pointers between structures. In these cases, "activation" is an abstract quantity, usually represented by a real number, that represents how much information processing activity is taking place on that structure. This is the interpretation usually given by psychologists (Collins & Loftus, 1975; Anderson, 1976), or by those computer representations of spreading activation (Fahlman, 1981; McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). In some cases, the network is interpreted more literally as being constructed out of physical nodes and interlinking relations (wires if the data base is an electronic circuit, or neurons if it is thought of as a neural network). In this case, activation is thought to be the actual electrical or chemical activity though the interconnections (e.g., see Feldman & Ballard, 1982).

Suppose one had a network representing the structure of animals (much as in Figure 7). How would a questions such as, "Does a shark have mass?" get answered? The spreading activation algorithm operates by starting at both "shark" and "mass" simultaneously. This activates the nodes for "shark" and "mass," which then, simultaneously, activate all of the relations that leave these two nodes. Activation spreads down the relations, taking time to do so, and reaches the nodes at the end of the relations. These nodes get activated and, in turn, spread activation down all the relations that lead from them. Imagine spreading rings of activation, each ring originating from one of the starting points. Eventually these expanding rings will coincide. When that happens, we know there is a path between the nodes that have originated the colliding rings of activation. That path can then be readily found by following the activation traces, and, depending upon the nature of the path, the question can then be answered.

---

[4] Actually, even the diagram illustrated in Figure *9D* is not quite accurate, for it shows the English names for the nodes and relations on the diagram. In fact, the names of the node and relations do appear within the network itself, but instead exist outside the network in what might be called "the vocabulary."

There are many details left out of this story. There are a large number of possible questions:

- How is the fact that two expanding rings of activation have intersected actually detected?
- How can the resulting path be followed?
- If there are *N* relations leaving a node, does the amount of activation depend upon *N?*
- Do the expanding rings of activation trace out all of the possible relations, or can they be restricted to a subset of the class of relations?
- For how long a period of time does activation leave a trace?
- Are there different kinds of activations? That is, is it possible to distinguish the activation left by one process from the activation left by another?
- What is the best possible way to model this process?
- What is the best possible way to construct a working, simulation model of this process?

These are the kinds of questions that have guided the research in this area. One of the major psychological issues addressed by activation studies has been the time course of activation (e.g., Maclean & Schulman, 1978; Neely, 1976). A second use of activation has been as a tool to examine the nature of the representation: if activation of one node will activate another, then the secondary activation "primes" any information processing that must make use of the other, thereby speeding its operation. Priming, therefore, is a technique that allows one to study the manner by which the interconnections are constructed. The basic priming study goes like this (after Meyer & Schvaneveldt, 1971): Subjects are asked to read two strings of letters and to decide as rapidly as possible whether each is a word or non-word. Thus, a typical pair of items might be "nurse plame." If the two words are related (as in "bread butter") the judgement that both are words is considerably faster than if the two are not related (as in "bread nurse"). The interpretation is that reading of the first word sends activation to words related to it, thus "priming" the other words and making their detection and judgement easier and faster. Clearly, this kind of result can be used to study the inter-relationships of items within memory by examining the amount of priming effect.

In a similar way, Collins and Quillian (1970) argued for support of their hierarchical organization of memory by demonstrating that prior exposure to the statement *A canary is a bird* reduced the amount of time that it took a person to determine whether it was true that *A canary can fly* more than it reduced the time to decide whether it was true that *A canary can sing*. They argued that to answer the question about flying, the node for "bird" had to be examined, and this was primed by the prior exposure, whereas to answer the question about singing, only the "canary" node was involved, and this was only minimally primed by the prior exposure.

Neely (1976) used priming as a technique to study Posner and Snyder's (1975) view of spreading activation. Posner and Snyder (1975) suggested that a visually presented word will automatically activate its representation, with the activation then spreading to the representations for other related words. This automatic activation is rapid, it occurs without attention or conscious awareness, and it has no effect upon unrelated items. Conscious activation can also occur, this time through the limited capacity, conscious-attention mechanism. This type of activation is slow, it requires attention and conscious awareness, and it can be applied to information unrelated to the item upon which it is focussed (usually by inhibiting these other items). The experimental procedure followed by Neely was to "prime" the subject by the presentation of a word, then, after a delay, to present a target item consisting of a letter string to the subject. The subject had to decide as quickly as possible whether or not the target item was a word. In some cases, the prime and the target were related, in other cases unrelated. In some cases the subject was told the relationship the target item would have to the prime, and in other cases the subject was not told. The critical test concerns what happens when the prime is a word like *building* and the test item a word like *door* or *arm*. When the subject thought that the word "building" would usually be followed by words that were parts of buildings, a facilitation on those words occurred, with no decrement in the ability to determine whether unrelated words, such as "arm," were words or not. Now suppose that the subject were told that whenever "building" occurred as the prime, the test word was likely to be a part of a body. In this case, the subject should activate "body" upon seeing the word "building." In fact, when the delay between the prime and the test item was short (less than 250 msec.), the results were essentially the same as in the first case: when the subject expected the prime of "building" to be followed by words that referred to parts of a building. However, when the delay was long (greater than 700 msec.), the speed to respond to body parts was increased and the speed to building parts decreased. Thus, it appears that spreading activation can be initiated either automatically, in which case it serves primarily to activate related concepts, or consciously, in which case it takes some time to be initiated, but it can both increase and inhibit the activation levels.

A third issue that has been widely investigated is whether or not the number of relations that leave a node affect the speed or amount of activation that goes down the interconnecting links. This is called the *fan effect*, and it has most widely been studied by Anderson and his collaborators. Anderson's model of cognition (ACT: Anderson, 1976) uses activation as one of its central themes, and so in addition to describing the fan effect that he has studied so extensively, let us also review the basic model.

*ACT and the fan effect*. ACT makes a set of processing assumptions that are used in conjunction with its representational assumptions (which are of the standard form we described for propositional representation) to make predictions about specific experiments. In particular, ACT consists of the following assumptions about memory structure.

(1) *Representation.* Information in memory is stored in network structures.

(2) *Activation.* Each node and each link in memory can be in one of two states, either *active* or not. The links connecting active nodes need not be active. If a link is active, the nodes it connects with become active; activation spreads from one node to the next through the active interconnecting links.

(3) *Strength of Links.* Each link has a strength *s* associated with it.

(4) *Spread of Activation: The fan effect.* The probability that activation will spread through a link is a function of the ratio of the strength of the particular link to the sum of the strengths of all of the links emanating from the node.

(5) *Active Lists.* Active nodes may be on an *active list*. The number of nodes that can be on the active list at one time is limited, but unless a node is on this list, its activity cannot be sustained for more than a short period.

Anderson assumes that the actual processing and interpretation is performed by an external interpreter that is in the form of a "production system" (more on this in a later section). The processor can put nodes on the active list (or remove them) and carry out the specific tasks required of the cognitive system as a whole.

One major set of investigations that have been motivated by the ACT system have been studies of the fan effect. Basically, the "fan" experiments are strong tests of assumption (4) and weaker tests of the other assumptions. In particular, the fan effect refers to the fact that the activation that goes across a link is inversely proportional to the number of links that "fan out" from or leave the node. This results in the somewhat non-intuitive prediction that the more one knows about something, the longer it takes to retrieve that information. This follows because the more links emanating from a particular node the longer, on average, it will take the activation to spread to adjacent nodes. Because the major mechanism for retrieving information makes use of the activation spreading along links, it should be possible to get rather direct information on the pattern of links from observations on retrieval time. The typical procedure for these experiments involves teaching subjects a set of facts arranged so that different numbers of facts apply to different concepts. In a typical experiment, experimental subjects are shown a number of sentences to learn and then tested on their ability to recognize test sentences. The results indicate that subjects are slower to recognize a sentence of the form "The doctor hated the lawyer" if they had learned other facts abut the lawyer and the doctor than if they had not. Thus, the more sentences of the form "The doctor loved the actor" and "the lawyer owned a Cadillac," the slower the recognition of the test sentence. The basic result is as predicted: the more facts, the slower the recognition time.

The basic "fan effect" might also be called "the paradox of the expert"; the theory appears to say that the more one knows about a topic, the slower will be the access to material about that topic. This flies in the face of common wisdom; could common wisdom be wrong? Smith, Adams, and Schorr (1978) challenged the result, pointing out that one difference between the knowledge structures of experts and the knowledge structures studied in these experiments is that we would expect the knowledge of experts to consist of a large amount of tightly inter-related structures, not just random facts like those in the basic fan experiment. Smith *et al*. tested this hypothesis by presenting their subjects with interrelated materials in which the facts about a specific topic formed thematic units. shows the materials from this experiment. Smith *et. al*. found, indeed, with these materials that the fan effect was greatly diminshed, and possibly reversed. In further studies, Reder and Anderson (1980) and Reder and Ross (1983) have shown that whether or not one gets a fan effect depends upon the exact question that must be answered by the subjects. When the subject must retrieve a particular proposition, the fan effect does indeed occur. However, when the same subject is asked to make a "consistency" judgement on the same information, the fan effect is reveresed; the more the subject knows about the item, the faster the response. Thus, there must be multiple processes acting upon the information within memory that yield different results for different tasks. Reder and Ross (1980) proposed that when subjects learn a consistent set of facts about a concept in memory, they generate sub-nodes upon which to attach the information. Without going into the details at this point, note that the theory makes a counter-intuitive prediction that appears to hold in appropriate circumstances, but that requires different processes to operate upon the same data structures within memory. The results again emphasize the fact that in studies of representation, it is not possible to separate the effects of the processes that operate upon the data structures from the data structures; the two must be considered together.

*Schank's conceptual dependency*. One of the more important applications of the semantic network has been the work of Schank and his colleagues on the representation of concepts (Schank, 1975, 1981; Schank & Abelson, 1977). Schank took seriously the task of creating a plausible representation of the kind of knowledge that underlies language use. He wanted a representation that was unambiguous and unique. He wished to be able to express the meaning of any sentence in any language. The representations were intended to be language independent; if two sentences had the same meaning, they should have the same representation whether they were paraphrases within a given language or translations between languages. Moreover, Schank wished concepts which were similar to have representations which were likewise similar. In order to carry out this process he proposed that all incoming information be stored in terms of a set of conceptual primitives. *Conceptual dependency theory*. was designed to interrelate these conceptual primitives in order to represent a wide range of different meanings. The first job with such an enterprise is to be very specific about what the representational primitives are and Schank, more than anyone, has taken this task seriously. He has proposed a list of eleven *primitive acts* which he believes underlie the representation of all concepts. These include five basic *physical* actions of people:

- **PROPEL** which means to apply force to;
- **MOVE** which means to move a body part;
- **INGEST** which means to take something inside of an animate object;
- **EXPEL** which means to take something that is inside an animate object and force it out;
- **GRASP** which means to grasp an object physically.

There are also two basic *change of state* acts:

- **PTRANS** (for physical transition) which means to change the location of something;
- **ATRANS** (for abstract transition) which means to change some abstract relationship (usually ownership) of an object.

Shank lists two *instrumental* acts:

- **SPEAK** which means to produce a sound;
- **ATTEND** which means to direct a sense organ towards some particular stimulus.

Finally, there are two basic *mental* acts:

- **MTRANS** (for mental transition) which means to transfer information such as from one person to another or from one part of the memory, say LTM (long term memory) to STM (short term memory);
- **MBUILD** (for mental build) which means to create or combine thoughts. This is involved in such concepts as thinking, deciding, etc.

In addition to these primitive acts, there are a number of other primitive elements which are combined to represent meanings. For example, there are **PPs** (picture producers) underlying the meanings of concrete nouns, sets of primitive states, such as **HEALTH, FEAR, ANGER, HUNGER, DISGUST, SURPRISE,** etc. There are also a set of *conceptual roles* which these various primitive elements can play such as **ACTOR, OBJECT, INSTRUMENT, RECIPIENT, DIRECTION,** etc. A simple example will suffice to illustrate how the various basic elements combine in Schank's representational system. Figure 10A shows the conceptual dependency representation for the sentence

(1) John gave Mary a book.

In this case, the verb "to give" has been represented as the primitive **ATRANS**, the

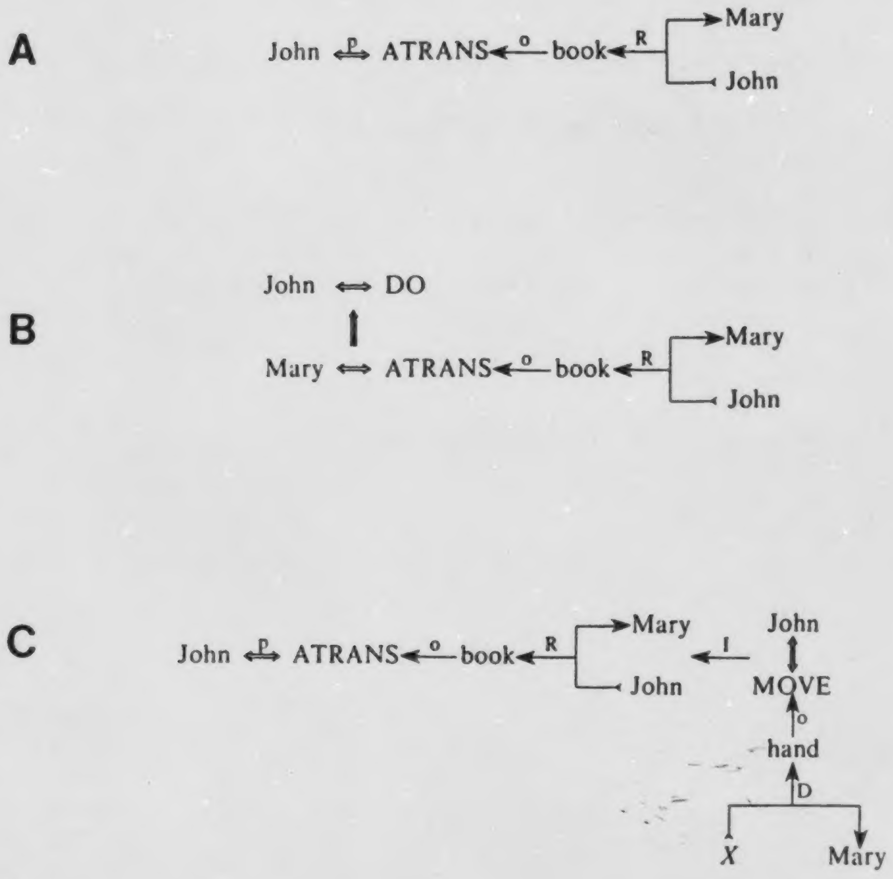

A
John
p
ATRANS
o
book
R
Mary
John
B
John
DO
Mary
ATRANS
o
book
R
Mary
John
C
John
p
ATRANS
o
book
R
Mary
John
I
John
MOVE
o
hand
D
X
Mary

Figure 10. The conceptual dependency representation underlying three interpretations of "John gave Mary a book." *A* shows the most basic interpretation of the sentence. *B* is the case in which John did something which allowed Mary to take the book and *C* shows the representation for John handing Mary the book. (From Schank, 1975, pp. 31-32.)

**ACTOR** (illustrated by the double arrow) is "John," the time is the past (illustrated by the p labeling the double arrow), the **OBJECT** is "book," and the **RECIPIENT** goes from "John" to "Mary." Note that the representation is *not* for the particular words of a sentence, but rather for the intended meanings. Thus, the figure represents only one interpretation of the sentence. The point is that in Schank's system, it is not sentences that have representations, rather it is meanings that are represented. Figure 10*A* represents the case in which John physically gave Mary the book. The same sentence could have been used for the case in which John had carried out some other action which let Mary take the book for herself. In this case, the correct representation would be the one illustrated in Figure 10*B*. Here, we see that "Mary" is now the **ACTOR** of the **ATRANS** and the action of "John" is the non-specific **DO**. Figure 10*B* illustrates the conceptual dependency underlying the case in which the same sentence means that John handed the book to Mary. In this case we see that "John" is again the **ACTOR** of the **ATRANS**, and that there is now an **INSTRUMENT** of the **ATRANS** specified. Note that the **INSTRUMENT** is an entire conceptualization which involves "John" MOVEing his hand from some location "X" to "Mary".

*KL-ONE*. In spite of their empirical successes, all of the various semantic network models have received various criticisms. In particular, Woods (1975) challenged the consistency and adequacy of these models to represent many of the distinctions of meaning that can be expressed in the predicate calculus and other logical formalisms. More recently, Brachman (1979) has furthered Wood's critique and proposed a new semantic network formalism, called KL-ONE (for "Knowledge Language One" and pronounced "clone"), that is intended to overcome the inadequacy of the previous models.

Woods and Brachman pointed out that the concepts of nodes and relations were imprecisely specified and inconsistently used. What exactly does it mean to connect one node to another with a labelled relation? What does a node or relation really stand for? Sometimes a node or a relation would stand for one kind of thing, other times for another. To begin, consider the nature of relations. Sometimes, as in Quillian's early work, a relation is treated as an attribute and the thing it points to as a value. Thus, a relation labeled COLOR might point from the node APPLE to the node RED. Other times, the relations might be labeled with transitive verbs and point from the subject to the object. Thus, the sentence that *The ball is on the table* is, according to some semantic network representations, characterized as a link labeled ON pointing from BALL to TABLE. More complex cases occur when three place predicate must be represented. Thus, the sentence *The ball is between the table and the chair* simply doesn't fit into the same format. Other semantic networks have links stand for still other things. In this case, some links, like *type* point from a token to a type. Other links, like *agent* or *recipient* do not stand on their own, but are only interpretable in the context of all of the other links on the node. Other links, like *iswhen* play still other special functions. The complaint is not so much that links are not used consistently but that so many different kinds of links are used to mean so many different kinds of things. Without a good deal of explication, it is easy to be confused about the meaning of a link. In particular, although all semantic network representations look superficially similar, a careful analysis of what the relations are actually used for and how they actually work shows that the similarity between systems and the homogeneity within systems is, at best, only superficial and, at worst, misleading and leads to errors.

Similar arguments apply to nodes. In particular, Woods argued that semantic network structures must represent the *intensions* of concepts. The term *intension* is to be contrasted with the term *extension*. The *extension* of a concept is the set of things that it denotes, whereas the *intension* of a concept is its internal structure, by virtue of which it denotes what it does. These correspond to what Frege (1892) called *Sinn (sense)* and *Bedeutung (reference)*. Concepts both refer to things (extensions) and have a sense (intension). Two concepts could both refer to the same thing in the world, but have different senses or intensions. A famous example of this is the contrast between Morning Star and Evening Star, both having the same extension (because they both refer to the planet Venus), but with each having a different intension.

Consider, as an example, the network structure illustrated in Figure 11*A*. Various semantic network theorists might wish to say that it represents the fact "John sees an airplane." Figure 11*B* might be said to represent the fact "John wants to see an airplane." Notice that the shaded part of Figure 11*B* is identical to the structure for Figure 11*A*, but the meaning of these two structures is different in the two cases. In the first case, we can conclude that there was an airplane that John saw: this would be an extensional interpretation. In the second, we can make no such interpretation. The node *airplane represents a real airplane in Figure 11*A*, but only a hypothetical one in 11*B*. Representational systems must distinguish between these two meanings of nodes --- the extensional and the intensional.

Brachman (1979) developed a semantic network type representational system designed to be very clear about the semantics of the networks. In particular, Brachman developed a system in which distinctions among the "type" classes of links were clearly marked and in which concepts were always intensional. Brachman called his kind of network a "structured inheritance net" (SI-Nets) and called his implementation of the idea KL-ONE.

There are two kinds of concepts in KL-ONE: *generic* and *individual*. Generic concepts represent classes of individuals; individual concepts represent particular individuals. Generic concepts represent classes by describing a prototype class member, organized in an inheritance hierarchy. Thus, as in traditional semantic networks, the concept for a term like "dog" might be represented as a specialization of the concept for a term like "animal."

Concepts themselves have an internal structure. The meaning of a given concept is determined jointly by its "superconcept" and its own internal structure. Internally, concepts consist of two major types of entities: *Role/Filler Descriptions* (roles) and *Structural Descriptions* (SD's). Every concept has a set of superconcepts, a set of roles which represent the conceptual components of the concept, and a set of SD's that describe the relationships among the various roles.

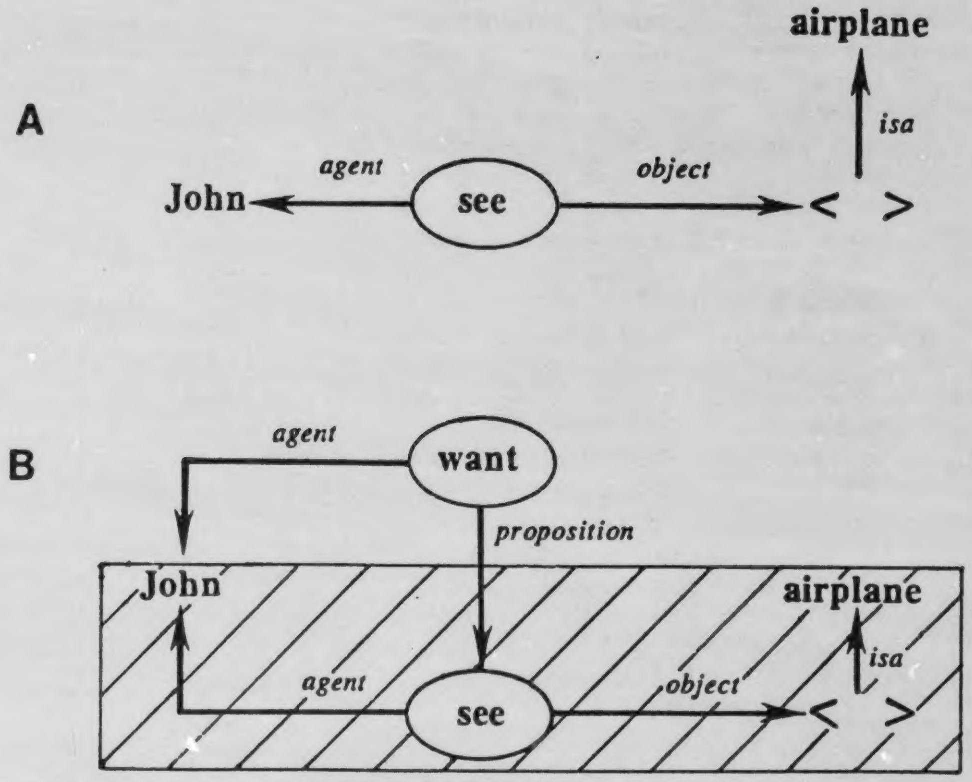

A
airplane
isa
agent
object
John
see
< >
B
agent
want
proposition
John
airplane
isa
agent
object
see
< >

Figure 11. The representational formalism adopted by KL-ONE (Brachman, 1979). *A*: "Johns sees an airplane." *B*: "John wants to see an airplane." The shaded portion of *B* is identical to *A*. the node <*airplane> represents a real airplane in *A*, but a hypothetical one in *B*.

Roles, too, have an internal structure. A given role has a *modality*, (is it an obligatory, optional, inherent or a derivable part?), a *Value Restriction (V/R)* (what kind of thing fills this slot?), a *RoleName* (an arbitrary name for internal reference only), and a *Number* (the number of such parts allowed for the particular concept). Figure 12 illustrates a KL-ONE representation of the concept *arch*. Arches have three roles, designated R1, R2 and R3. R1 represents the lintel or top of the arch. It is *obligatory*, in the network, it is locally called a "lintel," it must be a kind of "Wedge-Brick," and there can only be one of them in an arch. R2 represents the sides of the arch and is also obligatory, it is a kind of brick, it is locally called an "upright," and there are two of them. R3 represents the height of the arch. This is an *inherent* or *derivable* part: "vertical-clearance." The structural descriptions are the essential part of the concept: they indicate how the various parts are interconnected. Thus, for example, S1 gives the essential relationship between the UPRIGHTs and the LINTEL.

Knowledge in K-LONE is stored in strictly hierarchical structures. Thus, each KL-ONE concept is defined as a specialization of some higher level concept. In this definition, the relations between the roles of the concept and the superconcept must be specified, as must the relationship among the SD's of the concept and the superconcept. Figure 13*A* illustrates the relationship between a concept and its superconcept. Figure 13*B* shows that relationship for the case of an arch. In addition to the aspects of KL-ONE already discussed, KL-ONE has mechanisms for representing individual concepts and associated procedures.

In KL-ONE we have the latest and most sophisticated of the semantic network type representations. KL-ONE contains mechanisms for representing virtually all of the kinds of knowledge we have thus far described. It is, however, much farther from the empirical base than any of the other models.

### *Schemata and Frames*

So far, we have covered a variety of representational schemes that focus upon the basic, elementary levels of representation. The semantic feature approaches focussed almost exclusively on the representation of word meanings, the predicate calculus focussed on the kind of knowledge that could be expressed in a single sentence, and the semantic network and the conceptual dependency formalisms strived to include both lexical level and sentential level knowledge. The one thing that all these systems have in common is that they represent all knowledge in a single, uniform format. What is needed is the ability to introduce higher levels of structure. There is a need for representations which represent *supra-sentential knowledge*. In this case the goal is not to remedy the expressive problems of other representational methods, but to change the level of discourse.

The movement towards systems that focussed on higher units of knowledge was signaled by the publication, in 1975, of four papers: "A framework for representing knowledge" by Minsky, "Notes on a schema for stories" by Rumelhart, "The structure of episodes in memory" by Schank, and "Concepts for representing mundane reality in plans" by Abelson. Over the next several years, these papers led to the development of a number of related knowledge representation proposals, all aiming at the

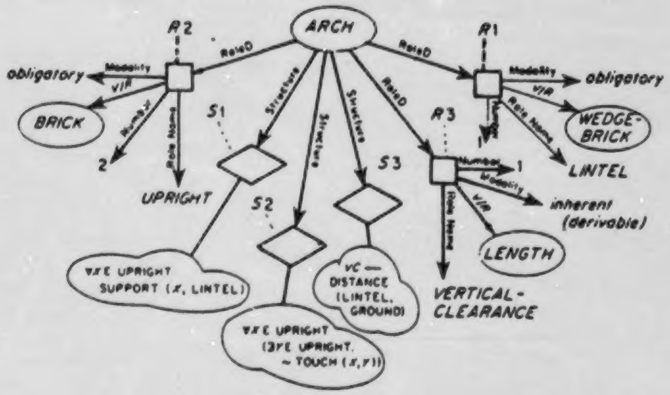

ARCH
R2
R1
R3
RoleD
RoleD
RoleD
Structure
Structure
Structure
obligatory
Modality
V/R
Number
Role Name
BRICK
2
UPRIGHT
S1
S2
S3
Modality
V/R
Number
Role Name
obligatory
WEDGE-BRICK
1
LINTEL
Number
1
Modality
inherent (derivable)
V/R
LENGTH
Role Name
VERTICAL-CLEARANCE
∀x∈ UPRIGHT SUPPORT (x, LINTEL)
∀x∈ UPRIGHT (∃y∈ UPRIGHT. ~ TOUCH (x, y))
VC — DISTANCE (LINTEL, GROUND)

Figure 12. A schematic representation of the KL-ONE representation for the concept of "ARCH." (From Brachman, 1979, p. 37.)

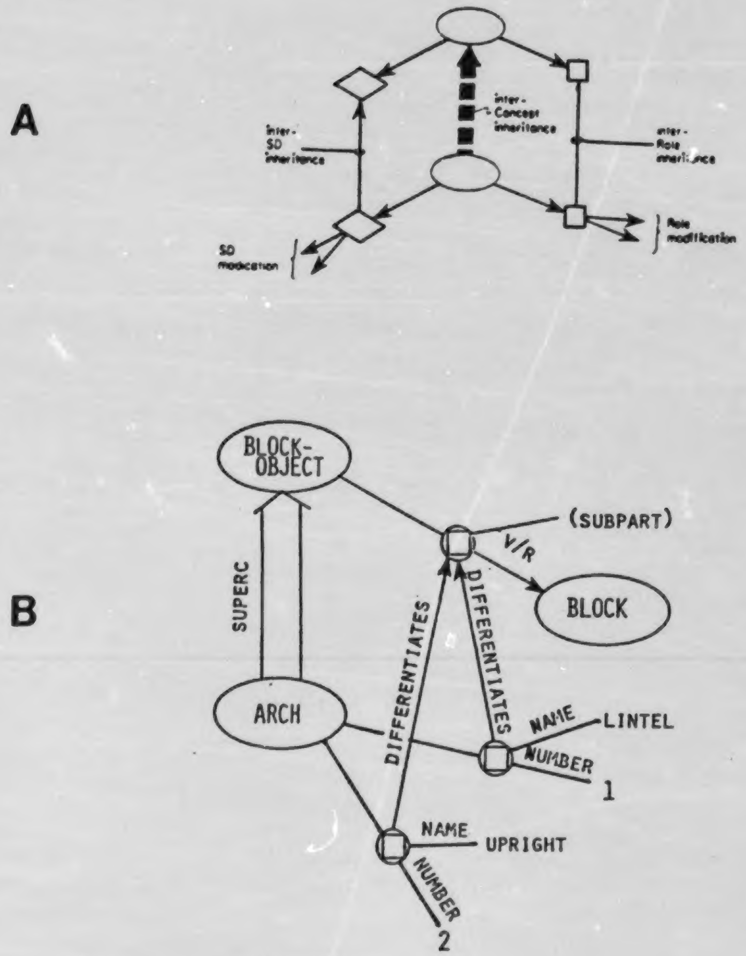

A
inter-
Concept
inheritance
inter-
SD
inheritance
inter-
Role
inheritance
SD
modication
Role
modification
B
BLOCK-
OBJECT
(SUBPART)
V/R
BLOCK
SUPERC
DIFFERENTIATES
DIFFERENTIATES
ARCH
NAME
LINTEL
NUMBER
1
NAME
UPRIGHT
NUMBER
2

Figure 13. *A:* The relationship between a concept and its superconcept, in KL-ONE (shown in schematic form). *B:* The relationship for the case of *arch*.

representation of suprasentential knowledge units. In his paper introducing the concept of the *frame* as a knowledge representation formalism, Minsky put the argument this way:

> It seems to me that the ingredients of most theories both in artificial intelligence and in psychology have been on the whole too minute, local, and unstructured to account -- either practically or phenomenologically -- for the effectiveness of common sense thought. The "chunks" of reasoning, language, memory, and "perception" ought to be larger and more structured, and their factual and procedural contents must be more intimately connected in order to explain the apparent power and speed of mental activities. (Minsky, 1975, p. 211)

A number of theorists have developed representational systems based on these "larger" units. We will discuss three of them here:

- A theory of schemata as developed by Rumelhart and Ortony (1977) and extended by Rumelhart and Norman (1978) and Rumelhart (1981).
- A theory of scripts and plans developed by Schank and Abelson (1977) and further elaborated into MOPS by Schank (1980).
- KRL, the first of the knowledge representation languages, developed by Bobrow and Winograd (1977).

The basic underlying feature of these theories is that the earlier work was useful in providing a foundation for further work, but that it was focussed on the wrong level to be useful in the understanding of understanding. The nodes and relations of semantic networks, the formulas of predicate calculus, and the feature lists of semantic concepts do have a place in the structure of representation, but they do not allow one to structure knowledge into higher-order representational units. The major function of these new approaches is to add such structure, wholistic units that allow for the encoding of more complex inter-relationships among the lower level units. These higher order units were given different names by each of the theorists: frame (Minsky), schema (Rumelhart & Norman), script (Schank & Abelson), and unit (Bobrow & Winograd). Nonetheless, the motivating force and in most cases the underlying themes are similar. We now turn to examine these higher-level structures.

*Summary of the major features of schemata*. The notion of the schema finds its way into modern cognitive psychology from the writings of Bartlett (1932) and from Piaget (1952). Throughout most of its history, the notion of the schema has been rejected by main stream experimental psychologists as being too vague. Recently, however, as we have begun to see how such ideas might actually work, the notion has become increasingly popular. In this section, we sketch the basic ideas of the schema, particularly as developed in the papers by Rumelhart and Ortony (1977), Bobrow and Norman (1975), Rumelhart and Norman (1978) and by Rumelhart (1981). For the most part, the characteristics of the schema as developed in these papers is consistent with the work of the other writers on the subject. However, as we will will indicate below, there are features which differentiate the ideas as well.

Schemata are data structures for representing the generic concepts stored in memory. There are schemata for generalized concepts underlying objects, situations, events, sequences of events, action and sequences of actions. interrelationships that is believed to hold among the constituents of the concept that it represents. Schemata in some sense represent the stereotypes of these concepts. Roughly, schemata are like models of the outside world. To process information with the use of a schema is to determine which model best fits the incoming information. Ultimately, consistent configurations of schemata are discovered which, in concert, offer the best account for the input. This configuration of schemata together constitutes the *interpretation* of the input. There appear to be a number of characteristics of schemata that are necessary (or at least useful) for developing a system that behaves in this way. Rumelhart (1981) and Rumelhart and Ortony (1977) listed several of the most important features of schemata. These include:

(1) Schemata have variables;

(2) Schemata can embed, one within another;

(3) Schemata represent knowledge at all levels of abstraction;

(4) Schemata represent *knowledge* rather than *definitions;*

(5) Schemata are active recognition devices whose processing is aimed at the evaluation of their goodness of fit to the data being processed.

Perhaps the central feature of schemata is that they are packets of information that contain *variables*. Roughly, a schema for any concept contains a fixed part, those characteristics which are always (or nearly always) true of exemplars of the concept, and a variable part. Thus, for example, the schema for the concept **DOG** would contain constant parts such as "a dog has four legs," and variable parts such as "a dog's color can be black, brown, white, . . ." Thus, **NUMBER-OF-LEGS** would be a constant in the schema, whereas **COLOR** and **SIZE** would be variables. Similar, in the **GIVE** schema the aspects involving a change of possession would be constants, and those aspects involving who the giver or the receiver was would be variables. There are two important aspects of variables for schema-based systems. In the first place, variables have *default values*. That is, the schema contains information about what values to assume for the variables when the incoming information is unspecified. Thus, consider as an example the following story sentences:

(2) Mary heard the ice cream truck coming down the street. She remembered her birthday money and rushed into the house.

In processing such a text, people usually invoke a schema for ice cream trucks going through a community selling ice cream to the children. In this schema there is a fixed part involving the relationships among the characters of the ice cream truck drama and a variable part concerning the particular individuals playing the particular roles in this drama. In this case, we tend to interpret Mary as the filler of the **BUYER** variable in the schema. Although the story tells us nothing about the age of Mary, we tend to think of her as a little girl. Thus, the default value of the age of the **BUYER** in this schema is childhood, and unless otherwise indicated, we tend to assume that this is the age of the **BUYER**. Default values can, of course, be overcome by explicit information in the incoming information. A second important aspects of variables involve our knowledge of the plausible range over which the fillers of a particular variable might vary. Thus, consider for example, the following examples:

(3) The child broke the window (with a hammer).

and

(4) The hammer broke the window (with a crash).

In the first case, we are likely to assign "the child" to the **AGENT** variable of the **BREAK** schema and to assign "hammer" to the **INSTRUMENT** variable. We might naively be tempted to assign "the hammer" to the **AGENT** role in the second example (after all child and hammer are both subjects of the verb). However, we know that hammers lie outside of the class of possible **AGENTs** for the schema and a much better fit is attained with the mapping of "hammer" onto the **INSTRUMENTal** variable in the second sentence as well. Thus, the process of interpretation involves the selecting of schemata to account for the input the the determination of which aspects of the incoming information map onto which variables of the schema. We say that the variables are *bound* to various parts of the incoming array of information. The binding of a variable involves assigning an interpretation to that part of the situation.

A second important characteristic of schemata is that they can embed one within another. Thus, in general, a schema consists of a configuration of sub-schemata. Each sub-schema in turn consists of configuration of sub-schemata, etc. Some schemata are assumed to be primitive and to be undecomposable. Thus, we might imagine that the schema for a human body consists, in part, of a particular configuration of a head, a trunk, two arms, and two legs. The schema for a head, contains, among other things, a face, two ears, etc. The schema for a face contains a particular configuration of two eyes, a nose, a mouth, etc. The schema for an eye contains an iris, an upper lid, a lower lid, etc. The schemata at the various levels can offer each other mutual support. Thus, whenever we find evidence for a face, we thereby have evidence for two eyes, a nose, and a mouth. We also have evidence for a head, and thereby, perhaps for an entire body. Thus, unlike the attribute or featural representational systems in which features are generally viewed as unitary elements, the schema theories propose a whole hierarchy of additional levels.

The third characteristic of schemata is that they represent knowledge at all levels of abstraction. Just as theories can be about the grand and the small, so schemata can represent knowledge at all levels -- from ideologies and cultural truths, to knowledge about what constitutes an appropriate sentence in our language, to knowledge about the meaning of a particular word, to knowledge about what patterns of excitations are associated with what letters of the alphabet. We have schemata to represent all levels of our experience, at all levels of abstraction. Thus, the schema theories suppose that the human memory system contains countless packets of knowledge. Each packet specifies a configuration of other packets (sub-schemata) which represent the constituents of the schema. Furthermore, these theories assume that these packets themselves vary in complexity and level of application.

The fourth characteristic involves the kinds of information that schemata are assumed to represent. We believe that schemata *are* our knowledge. All of our generic knowledge is embedded in schemata. When we think of representations for word meanings, we can imagine that we might wish to represent one of two kinds of information. One the one hand, it has been common for representational theorists to assume that word meanings are rather like what one might find in a dictionary --- the essential aspects of the word meanings. On the other hand, one might assume that the meaning of a word is represented by something more like an encyclopedic article on the topic. In this case one would expect that in a schema for a concept like "bird," we would have in addition to the dictionary knowledge, many facts and relationships about birds. A third kind of information needs to be represented: our experiences with birds. The first two kinds of knowledge about birds are referred to as *semantic memory*. The third kind of knowledge is referred to as *episodic memory* (the terms were invented by Tulving, 1972). It is generally assumed that schemata must exist for both semantic and episodic memory, and that schemata for semantic memory contain a great deal of world knowledge and are much more encyclopedic than dictionary-like.

Finally, schemata should be envisioned as active processes [5] in which each schema is a process evaluating its goodness of fit, binding its variables, and sending messages to other schemata that indicate its current estimate of how well it accounts for the current data. In this case, it is useful to distinguish between two data sources that a schema can use in evaluating its goodness of fit:

1. information provided by the schema's sub-schemata on how well they account for their parts of the input *(bottom-up information);*

2. information from those schemata of which the schema is a constituent about the degree of certainty that they are relevant to structuring the input *(top-down information).* The process of interpretation can consist of repeated processing loops as various schema interact with top-down and bottom-up information processing in an attempt to find the best overall fit. Eventually, the process settles down. The set of schemata that has the best goodness of fit to the input constitutes the final interpretation of the input data.

---

[5] Not all versions of schema theories emphasis this feature, but it is a useful conceptualization. See the discussion by Rumelhart (1981).

*Scripts, plans and MOPS.* According to schema theory the memory system consists of an enormous number of packets of knowledge. Schank, Abelson and their colleagues (c.f., Schank & Abelson, 1977) have developed specific examples of the knowledge one might have stored. This allows us to determine whether the system has practical value, that is, whether such knowledge could really serve as the basis for the kind of interpretations we get of stories we read. Schank and Abelson have developed a number of specific kinds of schemata, the simplest type being the *script*. A script can be thought of as a schema for a frequently occurring sequence of events. Schank and Abelson suggest that there are scripts for very common types of social events. For example, they suggest that there are scripts for a visit to a restaurant, for a visit to a doctor, for a trip on a train, and many other similar frequently occurring event sequences. The script which has received the most attention is that for the *restaurant*. Figure 14 gives Schank and Abelson's proposal for the restaurant script. A script, like all schemata, has a number of variables. These can be divided roughly into two categories, those which require a person to fill them (called *roles* ) and those which must be filled by objects of a certain kind (called *props* ). Each script contains a number of *entry conditions*, a sequence of *scenes*, and a set of *results*. Script processing, like schema processing in general, allows one to make inferences about aspects of the situation which were not explicitly mentioned. Consider the following example:

(5)

Mary went to a restaurant.
She ordered a quiche.
Finally, she paid the bill and left.

Once it is determined that the Restaurant script is the proper account for this little story, it is possible to make a large number of inferences. In the first place, we can assume that when Mary started the episode, she was hungry. We also can assume that she had some money before she went into the restaurant and that she ate the quiche before she paid the bill. We further assume that there was a waiter or waitress who brought her a menu, that she waited for the food to be served, and so on. Thus, among other things, the script provides the structure necessary to understand the temporal order of events. In communicating, we need only provide enough information to be certain that our listener finds the correct script, and we assume the rest follows automatically. The script itself allows the listener to infer many of the details.

Bower, Black and Turner (1979) carried out a number of experiments designed to evaluate the script as an explanation for how people actually understand and remember stories. Their first tack was to collect some direct evidence on the kinds of scripts that people in our culture actually have for such things as going to a restaurant, attending a lecture, going to a grocery store, getting up in the morning, and going to a physician. They then developed a composite script by assigning an importance to each action depending on how many students named that aspect. The results of this experiment are shown in Figure 15.

Bower, Black and Turner also looked for the expected inferences to show up when their subjects recalled stories. The procedure was to present a story in which only some of the events in the script were explicitly mentioned, then to see whether, THEORETICAL RESTAURANT SCRIPT (ADAPTED FROM SCHANK & ABELSON, 1977)

---

*Name: Restaurant*

*Props:* Tables
Menu
Food
Bill
Money
Tip

*Roles:* Customer
Waiter
Cook
Cashier
Owner

*Entry Conditions:* Customer hungry
Customer has money

*Results:* Customer has less money
Owner has more money
Customer is not hungry

---

*Scene 1: Entering*
Customer enters restaurant
Customer looks for table
Customer decides where to sit
Customer goes to table
Customer sits down

*Scene 2: Ordering*
Customer picks up menu
Customer looks at menu
Customer decides on food
Customer signals waitress
Waitress comes to table
Customer orders food
Waitress goes to cook
Waitress gives food order to cook
Cook prepares food

*Scene 3: Eating*
Cook gives food to waitress
Waitress brings food to customer
Customer eats food

*Scene 4: Exiting*
Waitress writes bill
Waitress goes over to customer
Waitress gives bill to customer
Customer gives tip to waitress
Customer goes to cashier
Customer gives money to cashier
Customer leaves restaurant

---

**Figure 14. The Restaurant Script. (From Bower, Black and Turner, 1979 p. 179; adapted from Schank and Abelson, 1977.)**

EMPIRICAL SCRIPT NORMS AT THREE AGREEMENT LEVELS

| GOING TO A RESTAURANT | ATTENDING A LECTURE | GETTING UP | GROCERY SHOPPING | VISITING A DOCTOR |
|---|---|---|---|---|
| Open door | ENTER ROOM | *Wake up* | ENTER STORE | *Enter office* |
| *Enter* | *Look for friends* | Turn off alarm | GET CART | CHECK IN WITH RECEPTIONIST |
| *Give reservation name* | FIND SEAT | Lie in bed | Take out list | SIT DOWN |
| Wait to be seated | SIT DOWN | Stretch | Look at list | Wait |
| Go to table | Settle belongings | GET UP | Go to first aisle | Look at other people |
| BE SEATED | TAKE OUT NOTEBOOK | Make bed | *Go up and down aisles* | READ MAGAZINE |
| *Order Drinks* | *Look at other students* | *Go to bathroom* | PICK OUT ITEMS | *Name called* |
| Put napkins on lap | *Talk* | Use toilet | Compare prices | Follow nurse |
| LOOK AT MENU | Look at professor | *Take shower* | Put items in cart | *Enter exam room* |
| *Discuss menu* | LISTEN TO PROFESSOR | *Wash face* | Get meat | Undress |
| ORDER MEAL | TAKE NOTES | Shave | Look for items forgotten | *Sit on table* |
| *Talk* | CHECK TIME | DRESS | Talk to other shoppers | Talk to nurse |
| Drink water | Ask questions | Go to kitchen | Go to checkout counters | NURSE TESTS |
| *Eat salad or soup* | Change position in seat | Fix breakfast | *Find fastest line* | Wait |
| Meal arrives | Daydream | EAT BREAKFAST | WAIT IN LINE | Doctor enters |
| EAT FOOD | Look at other students | BRUSH TEETH | *Put food on belt* | Doctor greets |
| Finish meal | Take more notes | Read paper | Read magazines | Talk to doctor about problem |
| *Order Desert* | *Close notebook* | *Comb hair* | WATCH CASHIER RING UP | Doctor asks questions |
| *Eat Desert* | *Gather belongings* | *Get books* | PAY CASHIER | DOCTOR EXAMINES |
| Ask for bill | Stand up | Look in mirror | *Watch bag boy* | Get dressed |
| Bill arrives | Talk | Get coat | Cart bags out | Get medicine |
| PAY BILL | LEAVE | LEAVE HOUSE | Load bags into car | Make another appointment |
| *Leave Tip* | | | LEAVE STORE | LEAVE OFFICE |
| Get Coats | | | | |
| LEAVE | | | | |

Items in all capital letters were mentioned by the most subjects, items in italics by fewer subjects, and items in small case letters by the fewest subjects

**Figure 15. Empirically determined scripts at three different levels of agreement. The events listed in all capital letters were the most frequently mentioned, those in italics the next most frequently mentioned items and those in lower case letters were mentioned by still fewer subjects. (From Bower, Black and Turner, 1979, p. 182.)**

in a subsequent recall, subjects recalled events that were part of the script, but not part of the material actually mentioned in the story. The results indicated that under some conditions, as much as 30 percent of the events subjects recall are events mentioned in the script, but not in the story itself. Clearly, the scripts are potent determiners of a subjects recall.

According to Schank and Abelson, the script is only the simplest of the schema-like knowledge structures. Clearly, not all situations that we wish to understand consists of a sequence of high frequency events. Often, the knowledge structures we have to bring to bear to get an interpretation of the situation must consist of more general and more abstract schemata. One important type of such an abstract schema is what Schank and Abelson have called the *plan*. Plans are formulated to satisfy specific motivations and goals. Future actions can be expected to involve attempts to attain these goals. Consider the following example:

(6)

> John knew that his wife's operation would be very expensive.
> There was always Uncle Harry . . .
> He reached for the suburban phone book.

Many people, when they encounter this story, assume that John wants to borrow money from Uncle Harry and that he is reaching for the phone book to find Uncle Harry's phone number to ask for the money. Now, we probably don't have a specific script for this particular activity. We do, however, probably know that when people are presented with problems, they attempt to solve them. Thus, having identified the problem in the story (the cost of the wife's operation) we expect to see some problem solving behavior on the part of the protagonist, so that we interpret further activity as an attempt to solve the problem. Moreover, we can assume that subgoals will be generated along the way, and that further activities will be generated toward the solution of the subgoal. In this case, the primary goal is to pay for the operation; the plan is to borrow money from Uncle Harry. Borrowing money involves contacting Uncle Harry, which in turn leads to the subgoal of calling on the telephone, which involves the further subgoal of discovering his phone number, and so on. Rumelhart (1975, 1977) and Wilensky (1978) have shown that many stories can be analyzed by means of problem solving.

In one of their experiments Bower, Black and Turner (1979) found that subjects sometimes recalled events which occurred in one script (say a dentist script) in If different scripts are entirely different data structures, there is no reason to suppose that events from similar scripts would be more often confused than events from quite different scripts. This result prompted Schank to revise the notion of the script so that scripts are not stored in memory as a simple sequence of events, but are derived at the time they are used from smaller, more fundamental data elements (Schank, 1980). Those elements which combine to form scripts, Schank calls MOPS. Thus, the doctor script is not a unitary element. Rather, it is derived from the interrelationship of such MOPS as the **fix-problem-MOP**, the **health-care-MOP**, the **professional-office-visit-MOP** and many other MOPS. Figure 16 illustrates the configuration of MOPS that Schank assumes might underlie the doctor script.

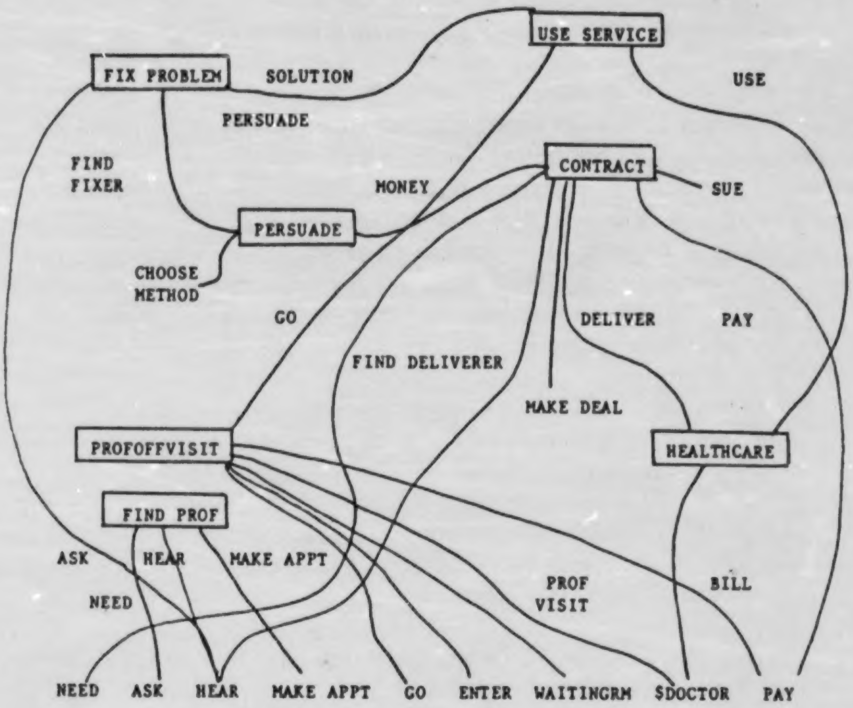

USE SERVICE
FIX PROBLEM
SOLUTION
USE
PERSUADE
FIND FIXER
CONTRACT
MONEY
SUE
PERSUADE
CHOOSE METHOD
GO
DELIVER
PAY
FIND DELIVERER
MAKE DEAL
PROFOFFVISIT
HEALTHCARE
FIND PROF
ASK
HEAR
MAKE APPT
NEED
PROF VISIT
BILL
NEED
ASK
HEAR
MAKE APPT
GO
ENTER
WAITINGRM
$DOCTOR
PAY

Figure 16. The configuration of MOPS which are assumed to underlie our knowledge of a doctors visit. (From Schank, 1980, p. 137.)

*KRL: A knowledge representation language.* Bobrow and Winograd (1977) developed a formal computational language for dealing with representational issues that they call KRL (for Knowledge Representation Language). Their goals were slightly different from those of the systems we have described, for in addition to their interest in expanding our understanding of representational issues, they also wished to emphasize the utility of developing a computational tool for those interested in the construction of computer models. Thus, they emphasized control processes and computational issues as well as representational issues. In addition, they developed several important conceptual notions, including the concepts of *descriptions*, *perspectives*, and of *procedural attachment*. Bobrow and Winograd described their goals this way:

> Much of the work in Artificial Intelligence has involved fleshing in bits and pieces of human knowledge structures, and we would like to provide a systematic framework in which they can be assembled. Someone who wishes to build a system for a particular task, or who wishes to develop theories of specific linguistic phenomena should be able to build on a base that includes well thought out structures at all levels. In providing a framework, we impose a kind of uniformity (at least in style) which is based upon our own intuitions about how knowledge is organized. We state our major intuitions here as a set of aphorisms . . . .
>
> 1. Knowledge should be organized around conceptual entities with associated descriptions and procedures.
> 2. A description must be able to represent partial knowledge about an entity and accommodate multiple descriptors which can describe the associated entity from different viewpoints.
> 3. An important method of description is comparison with a known entity, with further specification of the desired instance with respect to the prototype.
> 4. Reasoning is dominated by a process of recognition in which new objects and events are compared to stored sets of expected prototypes, and in which specialized reasoning strategies are keyed to these prototypes.
> 5. Intelligent programs will require multiple active processes with explicit user-provided scheduling and resource allocation heuristics.
> 6. Information should be clustered to reflect use in processes whose results are affected by resource limitation and differences in information accessibility.
> 7. A knowledge representation language must provide a flexible set of underlying tools, rather than embody specific commitments about either processing strategies or the representation of specific areas of knowledge. (Bobrow & Winograd, 1977, p. 5. Numbering of the seven "aphorisms" was not done in the original.)

The list of aphorisms reveals much of the common agreement about properties of higher order structures, by whatever name. Thus, aphorisms 1, 3, and 4 reflect general properties of schemata, things that we have already discussed. Aphorisms 2 and 3 introduce the notion of "description." Aphorism 2 is of special interest, for it introduces the notion of "perspectives," an important concept, one that we elaborate in a moment. Aphorisms 5, 6, and 7 reflect processing considerations, important for any useable system (including biological systems), but not relevant to the discussions of this chapter, so we will not elaborate upon them except to note that even when processing issues are not of prime concern, the tight relationship between representational structure and processing is evident in these three aphorisms: in general, one cannot ignore the processing structure when dealing with knowledge structure. To translate this into psychological terms: psychologists interested in psychological mechanisms and knowledge structures cannot ignore the issues and constraints placed upon the human system by neurological structures.

*Descriptions* were introduced into KRL both as an important processing and representational structure and also from considerations of processes that might operate within human memory (Bobrow & Norman, 1975; Norman & Bobrow, 1979). The major issue concerns just how one should refer to a concept or record in memory. There are only a few possibilities;

- Give each record a unique name; refer to the record by means of that name. This corresponds to the use of Proper Names in language and such unique identifiers as catalog number, part number, employee number, or Social Security number.

- Put each record in a unique place; refer to the record by referring to the place. This corresponds to the use of street addresses, telephone numbers, and memory addresses in computer systems.

- "Point" at the desired record, much as the arrows in a semantic network point to the nodes to which the relations refer. This corresponds to the use of wires in electronic circuits to interconnect the parts of the circuit, or the wires in a telephone switchboard, through which one physically makes the desired connection.

Further discussion of these issues takes us away from our topic (but see Norman & Bobrow, 1979; Norman, 1982, pp. 37-44). Note that all of the representational systems we have examined so far use either the methods of unique names or of pointers to refer to their items. But what if you know neither the name nor the location (address) of the item to which you wish to refer? What if the memory structure does not make available unique addresses or pointers, nor readily make available unique names (which is what we suspect is true of human memory)? How then does one describe the item one is seeking? For KRL, Bobrow and Winograd suggest the use of descriptions (much as Norman and Bobrow suggest for human memory in general). *Descriptions* offer an alternative method of referring to the desired record by describing the item being sought.

Descriptions have several virtues aside from their ability to refer to other items. Perhaps the most important is that of *partial specification* in which it is possible to describe the characteristics that one knows of an item, without fully specifying the item. Essentially, this is what one requests from an eye witness to a crime, for example:

| | |
|---|---|
| Query; | *What did the criminal look like?* |
| Reply; | *It was a woman, very tall, with red hair.* |

The reply in this example is a description that partially specifies the person. It is not enough to identify the person uniquely, but it goes a long way to constrain the set of possibilities. In many cases, it might even be sufficient to yield a unique identification. Examples of the use of descriptors of this sort from KRL include:

- The specification for the last name of a person as:

  *{(a ForeignName)*
  *(a String with firstCharacter = "M")}*

- The specification for the husband of Mary as:

  *(the maleParent from (a Family with femaleParent = Mary))*

Descriptions are quite useful in specifying *default values*. In our earlier examinations of default values, we only looked at simple values. Consider, though, a default value constructed of a description of the sort used above: "a person with red hair, whose height is more than 6 feet." This clearly enhances the power of defaults, for it allows them to use a variable amount of power, sometimes specifying uniquely what exact thing is to serve as the default, sometimes, being able simply to specify the characteristics loosely and imprecisely.

The second major innovation of KRL was the development of *perspectives*. The basic notion is that the very same concept or event can often be viewed for different purposes, with different information desired with each viewing. Each of these views is called a "perspective." Thus, a restaurant may be viewed as a place to eat, in which case the type, quality, and cost of the food being served is of importance. But a restaurant might also be viewed as a commercial business (by a potential investor, for example), in which case it is the location, size, clientele, and balance sheet that are of importance. Which of these views is provided the system user depends upon which perspective is requested.

The mechanism for handling perspectives is always to describe an entity by comparing it with some other entity in the memory: this is the aphorism 3 of KRL, from the previous list. Bobrow and Winograd describe this property this way:

> The object being used as a basis for comparison (which we call the *prototype)* provides a *perspective* from which to view the object being described. The details of the comparison can be thought of as *further specification* of the prototype. Viewed very abstractly, this is a commitment to a *wholistic* as opposed to a *reductionistic* view of representation. It is quite possible (and we believe natural) for an object to be represented in a knowledge system only through a set of such comparisons. There would be no simple sense in which the system contained a "definition" of the object, or a complete description in terms of its structure .... This represents a fundamental difference in spirit between between the KRL notion of representation, and standard logical representation based on formulas built out of primitive predicates.
>
> In describing an object by comparison, the standard for reference is often not a specific individual, but a stereotypical individual which represents the *typical* member of a class. Such a prototype has a description which may be true of no one member of the class, but combines the *default* knowledge applied to members of the class in the absence of specific information. The default knowledge can itself be in the form of *intensional* description (for example, the prototypical family has "two or three" children) and can be stated in terms of other prototypes. (Bobrow & Winograd, 1977, pp. 7-8).

*Procedural attachment*, provided a means for active processes to be triggered by the knowledge structures (Bobrow & Winograd, 1977; Winograd, 1975). Procedures can be attached to KRL structures in much the same way that general information about the object is attached (e.g., that Mary is person). Procedures are of two forms: servants or demons. Servants are called when needed to perform some particular action (a typical servant resides on a "slot" labelled "to fill," meaning that when it is desired to fill the particular slot, then the servant procedure that resides there is the relevant one to use). Demons, when activated, await some special condition that causes them to do their actions. Thus, if a set of units about a person are being established, several demons may be activated, each looking for information relevant to the slot from which it was invoked. Suppose we had established a unit for a person, but did not know the person's name. If in the course of the ensuing interaction the person's name got invoked, the name demon would immediately see it and place a copy on the relevant structure within the relevant unit. Demons provide a powerful tool, for they allow general processing to continue while they sit alert for information relevant to themselves.

Although KRL represents an important contribution to the development of knowledge representation systems, in fact, KRL itself has not been used much. Rather, its importance has been in the exploration of a variety of representational issues. Most of the innovations of KRL such as descriptions, perspectives, and procedural attachments are now considered standard tools.

### *The Relationship of These Representations to Classical Associations*

Before we leave the discussion of Propositional Representation, it is useful to note the relationship between the representational systems described here and classical Association Theory. After all, are not these systems simply systematic presentations of the associations that everyone has long believed must exist among different items within memory? The answer is "yes, but no." Current representational models do indeed represent a formalization of associations. However, this is a new association theory: a neo-associationism. The basic propositional and procedural representation system contains pointers from one item within memory to another; these pointers correspond to the associations of the classic theory. However, these modern theories of representation -- especially Propositional and Procedural Representations -- differ from classic associations in four ways:

- The relations are *directed*. This means that the direction of the association matters, so that the association from A to B is not necessarily the same as that from B to A (and in general, is not the same). Some classical theories of Association had this property.

- The associations are *labelled*. This means that two items A and B can be associated in many different ways, and in following these associations heavy use is made of the differences among labels. The labels are meaningful, and different labels imply different logical relationships.

- A distinction is made between *types* and *tokens*. This overcomes one of the major problems of association theory in allowing a particular instance of an item to be activated without confusing it with all instances of the same item, or with the generic item itself.

- There is a distinction made among *levels of representation*. This allows for processing of higher order structures. Classical association theory (as well as early semantic networks, predicate-calculus, and set-theoretic representations) suffered from a homogeneity of representational levels, thus considerably weakening their power and inferential ability.

These four properties yield several important benefits, including enhanced powers of logical inference, including inheritance properties and a natural representation for default values. The distinction among levels of representation allows for the use of prototypical or generic units that can guide in the construction of new units or in the interpretation of existing ones. All in all, these properties enhance the powers of these neo-associational representations sufficiently well to overcome all the classic objections to them, as well as to solve some issues that were not even considered earlier. An excellent treatment of the relationship of semantic networks to association theory is given in the first section of the book by Anderson and Bower (1973).

## ANALOGICAL REPRESENTATIONS

Most of the representational systems we have discussed thus far were designed to represent information stored in long term memory. In particular, they were designed to represent *meanings*, which led naturally to propositional representations. But other considerations lead to other classes of representational ideas. Consider the representation of an *image*; how would one represent objects undergoing various transformations? A number of researchers, especially Shepard, Kosslyn, and their colleagues (c.f. Shepard & Cooper, 1982; Kosslyn, 1980), have proposed that the knowledge underlying images is *analogical* rather than *propositional*. There has been a good deal of debate concerning the nature of analog representations and of how they differ from propositional ones. In this section, we proceed by summarizing the work carried out under the rubric of analogical representations. We enter into the debate only after we have presented both points of view.

### *Shepard*

Shepard and his co-workers have focused primarily on a set of simple mental transformations of mental images. Most of their work has focused on a study of mental rotations. The general procedure is to present a picture of a pair of objects that either are similar or mirror images of one another, but that differ in orientation (see Figure 17). The subject's task is to decide, as quickly as possible whether the objects can be rotated into congruence. Typical data from these experiments are illustrated in Figure 18: the time to respond increases linearly and continuously as the angular difference between the two objects increase, whether they differ in picture plane orientation or in orientation in depth. Subjects often report that they do the task by imagining one of the objects being rotated into congruence with the other.

Based on their experimental findings, Metzler and Shepard (1974) argued that the process of mentally rotating an object involves the use of a *mental analog* of a physical rotation. There are, they argue, two characteristics of such an "analog process." First, an analog process

> has something important in common with the internal process that would go on if the subject were actually to perceive the one external object physically rotating into congruence with the other.

Second, in an analog process

> the internal representation passes through a certain trajectory of intermediate states each of which has a one-to-one correspondence to an intermediate stage of an external physical rotation of the object.
>
> . . .
>
> To speak of it [a process] as an analog type of process is . . . to contrast it with any other type of process (such as feature search, symbol manipulation, verbal analysis, or other "digital computation") in which the intermediate stages of the process have no sort of one-to-one correspondence to

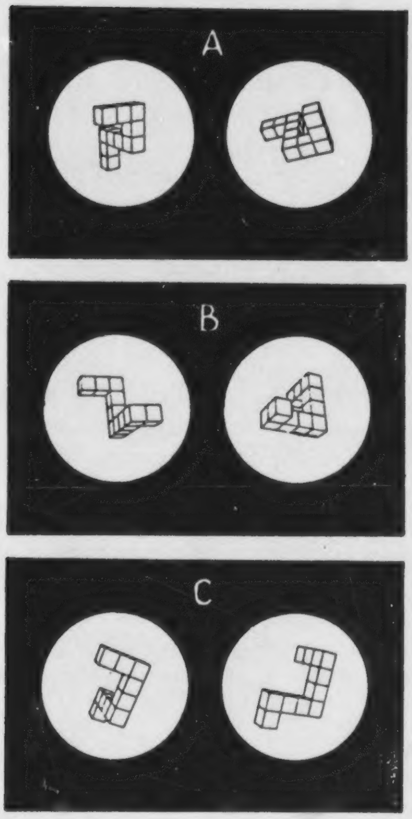


Figure 17. Illustrative pairs of perspective views, including a pair differing by an 80° rotation in the picture plane (*A*), a pair differing by an 80° rotation in depth (*B*), and a pair differing by a reflection as well as rotation (*C*). (From Metzler and Shepard, 1974.)

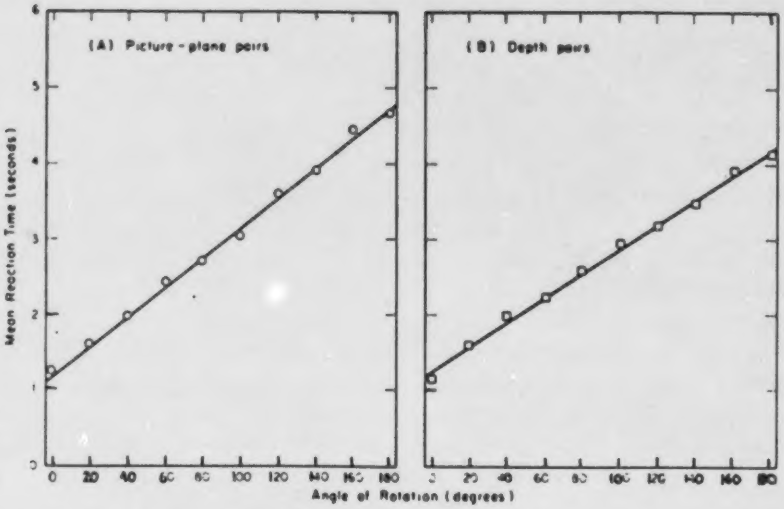


Figure 18. Mean time to determine that two objects have the same three-dimensional shape as a function of the angular difference in their portrayed orientations. (From Metzler and Shepard, 1974.)

> intermediate situations in the external world. (From Metzler & Shepard, 1974, pp. 150-151.)

In addition to the claim that the *processes* are analog, Shepard and his colleagues have argued that the *representations* themselves are analog: "The internal representation undergoing the rotation is viewed as preserving some degree of the spatial structure of its corresponding external object" (Cooper & Podgorny, 1976) and in this sense is an *analog* to the object itself (see also Shepard & Cooper, 1982, pp. 12-13).

The fact that the time to rotate something mentally grows linearly with angular difference does not, of course, mean that the process of mental rotation passes through the intermediate states. This datum by itself merely indicates that it takes longer to make the judgements the greater the angular disparity. In a very clever and important experiment, Cooper (1976) demonstrated that during mental rotation the internal representations do indeed pass through intermediate points and are, in that sense, analog. Subjects were to imagine an object rotating on a blank circular field. While they were doing this, a test object was presented in one of twelve orientations. The subject was to decide as quickly as possible whether it was the same as or a mirror image of the object being imagined. The critical feature of this experiment is that Cooper had previously determined the rate of mental rotation for her subjects, and therefore, depending on the initial orientation of the object being mentally rotated and on the time since the subject began, she could calculate the current orientation of the imagined object. Thus, she knew the angular difference of the test object and the imagined rotating object. The results, illustrated in Figure 19 showed that the greater the angular departure of the test stimulus from the orientation of the imagined stimulus, the longer it took the subjects to respond. It appears that subjects indeed form images of the object and that rotation involves the representation passing through intermediate orientations.

Despite the clarity of the empirical results, not everyone has been convinced of the need for an *analog* as opposed to a *propositional* representational system. There are three reasons for this. First, it is possible that a "propositional" system could be constructed which would produce the same results. Second, the kind of analog system envisioned by Shepard and his colleagues is clearly a special case system: it is not at all clear how it might interface with the kinds of propositional representational systems that have been so powerful in other domains. Third, it is not at all clear what the analogical system would look like in detail. How should these analog systems be represented in our theories? What would such a system actually look like? In what ways would it really be different from the representational systems we have discussed thus far? These questions have been addressed and tentative answers have been proposed by Kosslyn and his colleagues, and so we turn now to a discussion of this work.

### *Kosslyn*

The best articulated theory of image representation was put forth by Kosslyn and Schwartz (1978) and refined in Kosslyn (1980). Kosslyn's theory was built around what he called the Cathode Ray Tube (CRT) metaphor for visual imagery. Figure 20 illustrates the basic aspects of the metaphor. The basic idea is that there are two

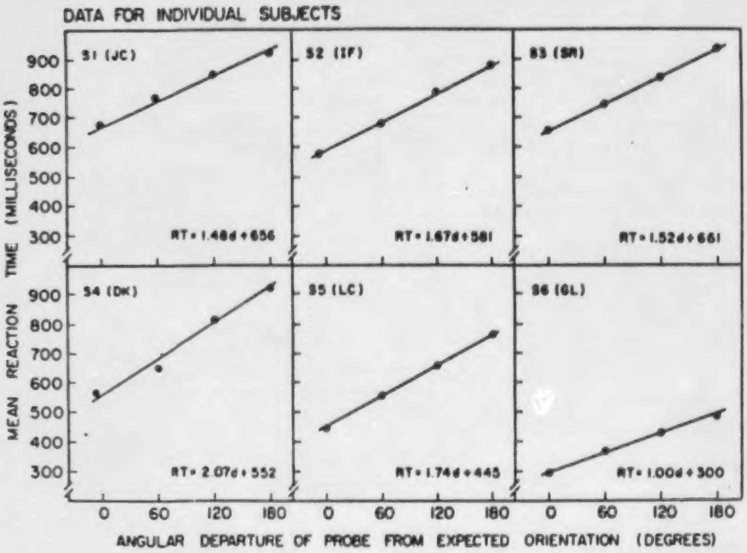


**Figure 19.** Mean reaction time to unexpected test probes, plotted as a function of angular departure of the test probe from the expected orientation, for each of the six individual subjects. (From Cooper, 1976, p. 168.)

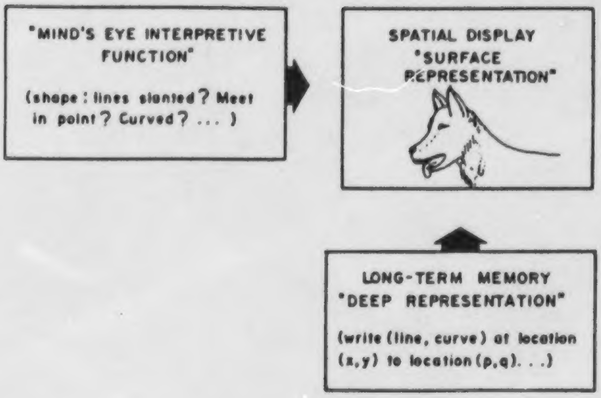

"MIND'S EYE INTERPRETIVE FUNCTION"
(shape : lines slanted ? Meet in point ? Curved ? ... )
SPATIAL DISPLAY "SURFACE REPRESENTATION"
LONG-TERM MEMORY "DEEP REPRESENTATION"
(write (line, curve) at location (x,y) to location (p,q). . .)

Figure 20. A schematic representation of the cathode-ray-tube (CRT) metaphor. (From Kosslyn, 1980, p. 6.)

fundamental kinds of representations of imaginal information. First, there is the *surface representation* corresponding to the visual image itself. This representation is assumed to occur in a "spatial medium" which imposes a number of characteristics on the image:

- Parts of the image represent corresponding parts of the imaged object in such a way that, for example, distance between parts of the representation correspond to distance between parts of the imaged object.
- Just as a CRT has a limited spatial extent, so an image should have a limited spatial extent: images that are too large can not be represented without overflowing.
- Surface representations of images, like those of CRTs, are assumed to have a "grain size," so that there is a loss of detail when an object is imaged too small.
- Images, like CRT screens, require a periodic "refreshing" without which they will fade away.

In addition to the surface representation, the CRT metaphor suggests that there is a *deep representation* from which the image is being generated. Kosslyn (1980) suggests that images are generated from some sort of propositional representation, so that the underlying memory representation may not have the same spatial properties as the surface image. The third suggestive aspect of the CRT metaphor involves the existence of an *interpreter* or "mind's eye" that processes the surface image and serves as an interface between the surface image and a more abstract "semantic" interpretation of the constructed image. The interpretive processes might involve some of the same processing mechanisms used in general visual processing.

Kosslyn has constructed a computer simulation model that offers plausible accounts of a variety of data on visual imagery. In his model, Kosslyn proposes that the *surface* representation consists of a matrix of points. An image is represented in the matrix by filling in the cells of the matrix.[6] The matrix is of limited *extent*, thus limiting how large an image can be; it has a particular *grain size*, thus limiting how small an image can be and still be seen clearly, and the matrix is organized so that the grain of the central region is smaller than the grain size of the peripheral region (the cells in the outer region of the matrix are not all used). Further, Kosslyn assumes that the representations in the visual matrix *fade* unless the old material is "refreshed" periodically. This is implemented by having the magnitude of the value within each cell of the matrix decrease with time after having been written into the matrix.

---

6. In computer graphics, this is known as a "bit map" representation.

As with the CRT model, the images in the computer model are not long term representations, but simply temporary representations that are constructed to aid in the solution of particular problems. The long term representations or *deep* representations contain the knowledge that allows the construction of the images. Consequently, Kosslyn has two kinds of long term representations. He uses a relatively standard *propositional* representation for storing general knowledge and also what he calls a *literal* representations for storing the data necessary to create an image. These literal images are themselves stored as a set of polar coordinates ($r,\theta$ pairs) with respect to an origin. The polar coordinates allow easy shifting of location of the image (by changing the origin), easy change of size (by multiplying the values of $r$ by a constant), and easy rotation (around its origin). Figure 21 shows the long term memory representations and the major processes of the theory.

There are three major classes of processes proposed by Kosslyn. These are **IMAGE, LOOKFOR** and various **TRANSFORMATIONS**. **IMAGE** is a procedure for generating an image from the stored representation. It constructs a whole image out of the literal representations of their parts and their descriptions. **LOOKFOR** scans the image, using the surface representation along with the long term memory description of the object and finds the location of the looked for object in the image, if it is in the image. There are also three image transformation operations: **SCAN, ZOOM, PAN** and **ROTATE**. **SCAN** moves the image within the matrix. **ZOOM** moves all points out from the center, leaving a larger image. **PAN** moves all of the points toward the center, creating a smaller image. **ROTATE** moves all points of an image around a pivot, thus rotating the surface image. All of these transformations operate in small steps so that the surface matrix goes through intermediate points as it processes. Thus, in Shepard's sense, Kosslyn's system is truly an analogical system.

Kosslyn has arrayed an impressive amount of evidence for many of the detailed assumptions of his theory. In one such experiment, Kosslyn, Ball and Reiser (1978) showed that the time to scan between two points on an image were proportional to the distance between those two points on the object being imaged. Thus, subjects were presented with a picture of a map (Figure 22*A*) and were asked to memorize it, particularly noticing the seven *X*'s on the seven key locations of the map. The subjects continued to study the map until they could reproduce it with great accuracy. They were then instructed to image the map and told to mentally stare at a named location. They were then given another location name and told to mentally scan to that location and press a button when they reached it. Figure 22*B* shows the results. Clearly, "mental scanning" depends on the "mental distance" over which the scan takes place.

In another experiment, Kosslyn (1975) showed that the time that it takes to verify that an image of an animal has a particular property depends on the imaged size of the animal. Thus, subjects were told to image a particular animal to be one of four relative sizes. The largest size was as large as they could imagine without "overflowing" their image, the others to be scaled down by a factor of six in each case. Subjects were then asked whether the image of the animal had a particular property (i.e., they were asked to image a rabbit and then asked whether a rabbit has claws). The time to answer the question depended strongly on the size of the image and not on the size of the animal. Figure 23 shows the results of this experiment.

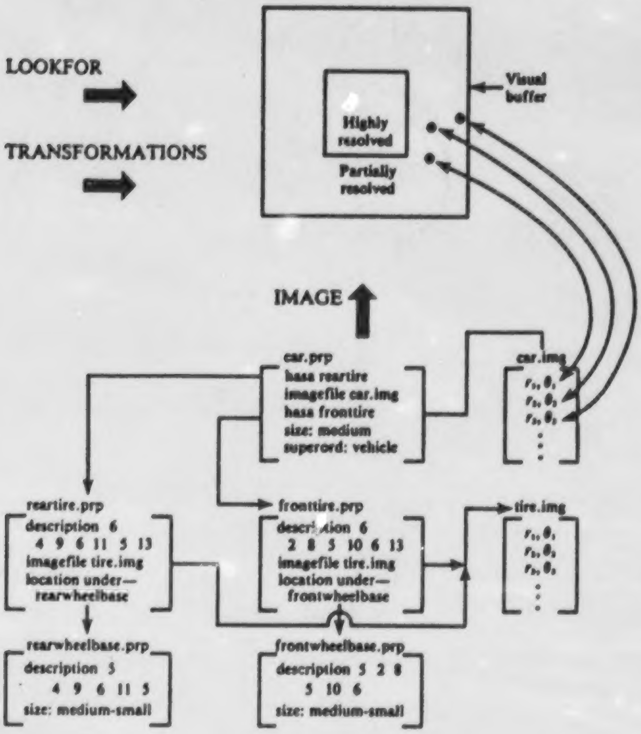

LOOKFOR
TRANSFORMATIONS
Visual buffer
Highly resolved
Partially resolved
IMAGE
car.prp
hasa reartire
imagefile car.img
hasa fronttire
size: medium
superord: vehicle
car.img
$r_1, \theta_1$
$r_2, \theta_2$
$r_3, \theta_3$
reartire.prp
description 6
4 9 6 11 5 13
imagefile tire.img
location under—
rearwheelbase
fronttire.prp
description 6
2 8 5 10 6 13
imagefile tire.img
location under—
frontwheelbase
tire.img
$r_1, \theta_1$
$r_2, \theta_2$
$r_3, \theta_3$
rearwheelbase.prp
description 5
4 9 6 11 5
size: medium-small
frontwheelbase.prp
description 5 2 8
5 10 6
size: medium-small

Figure 21. A schematic representation of the structures posited by Kosslyn. The major processes of the model **LOOKFOR** things in the image, perform **TRANSFORMATIONS** on the images and create an **IMAGE** from a long term memory representation. (From Kosslyn, 1980, p. 147.)

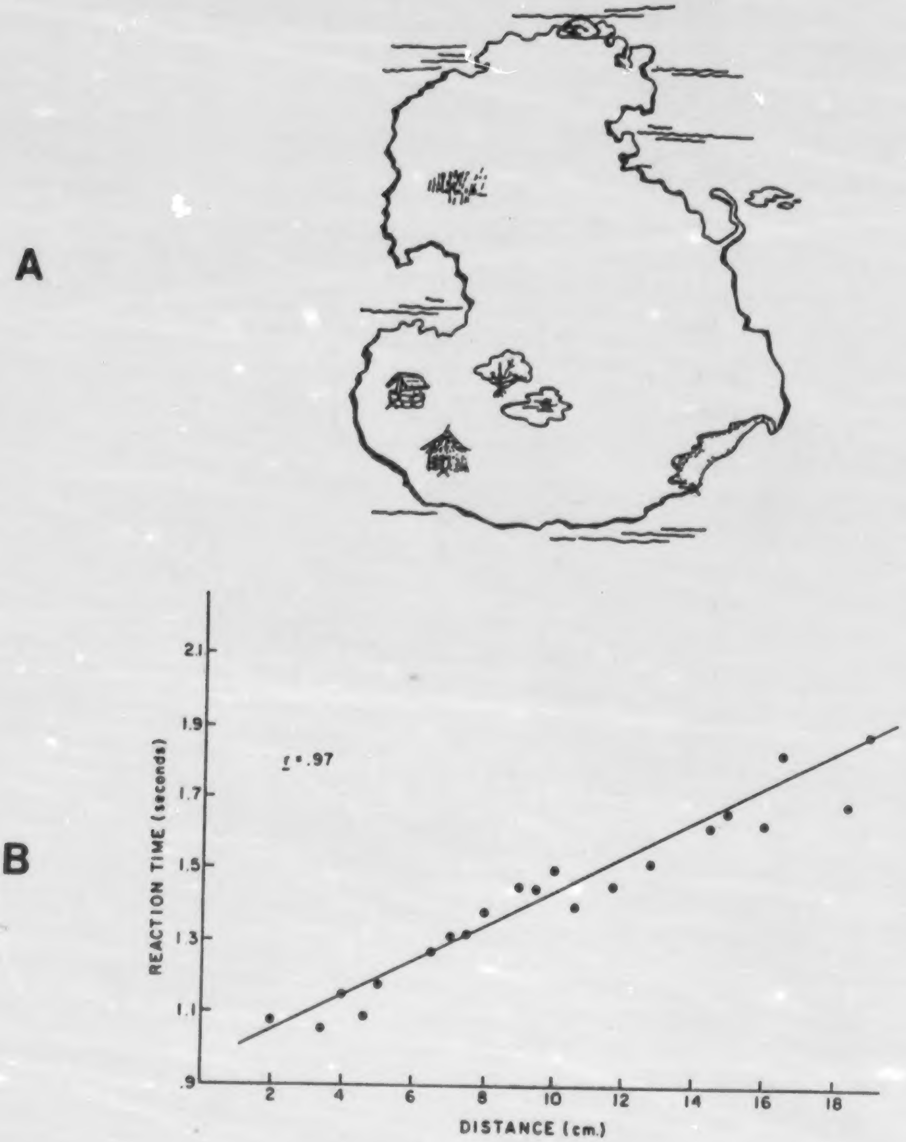


Figure 22. *A:* A fictional map subjects memorized and later imaged and scanned across. (From Kosslyn, 1980, p. 43.) *B:* The time to scan between all pairs of locations on an image of the map illustrated in *A*. (From Kosslyn, 1980, p. 44.)

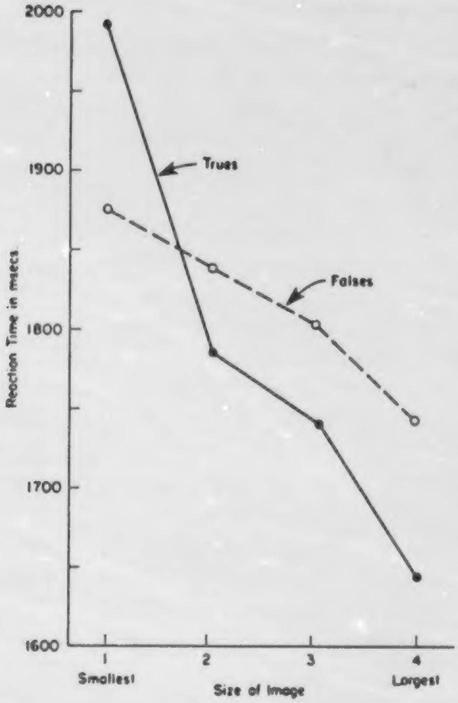


Figure 23. The time required to evaluate properties of animals imaged at one of four relative sizes. The largest size was to be as large as possible without overflowing and the rest were scaled down according to a training procedure. (From Kosslyn, 1980, p. 59.)

One of the important assumptions of Kosslyn's model is that the medium in which images are created has size limitations so that it will hold only a certain amount of material. Kosslyn wished to get an empirical measure of the size of the visual image or, as he called it, "the visual angle of the mind's eye." To do this, Kosslyn devised a "mental walk task" to measure the visual image. In these experiments people are asked to image particular objects as if the object were at some distance. They were then asked to mentally walk toward the object until it completely filled their mental image and to estimate the "mental distance" to the object. Using a variety of different imaged objects, Kosslyn found that the estimated distance at which a particular object was imagined to "overflow" the image was linearly related to the size of the object. Figure 24 shows the results for imagined line drawings of animals. These results suggest that the "visual angle" of the mental image subtends about 20°. Similar results were found for several other sets of imagined stimuli. Clearly, a visual image seems to have a definite perceived size and there seems to be substantial agreement about what that size is.

In addition to these results, Kosslyn has found that the time to create an image depends on the number of objects in an image, that an image of a large object takes longer to create than an image of a smaller object, that the fields on which visual images occur are roughly circular, and a number of other similar results. Based on these results, Kosslyn argues that the key features of the CRT model and its computer simulation are confirmed. In particular, Kosslyn (1980) argues that:

(1) Images occur in a spatial medium in which locations are accessed in such a way that the interval properties of physical space are preserved such that each portion of the image corresponds to a portion of the object being imaged. Evidence for this comes from introspection and from the results of the scanning experiments. Since the time to scan an image from one point to another is proportional to the actual distance between the points on the physical object being imaged, the image must be preserving the distance relations of the object.

(2) Images have a finite grain size. Evidence for this assumption comes from the experiments involved with judging properties of objects imaged at different sizes. The fact that parts of smaller objects are more difficult to "see", implies that things lose precision when they get too small in an image This precision is presumably determined by the grain size of the imaginal medium.

(3) The imaginal medium has a definite size and shape which limits the amount that can be imaged at one time. Evidence for this comes the experiments involving the size of the "visual angle of the mind's eye." Since images of large objects overflow the image at greater subjective distances than images of smaller objects it appears that the size of the imaginal medium is a limiting factor on the size of the image. Similarly, since the subjective distance at which a ruler overflows is independent of the imaged orientation of the ruler, the medium must be roughly circular.

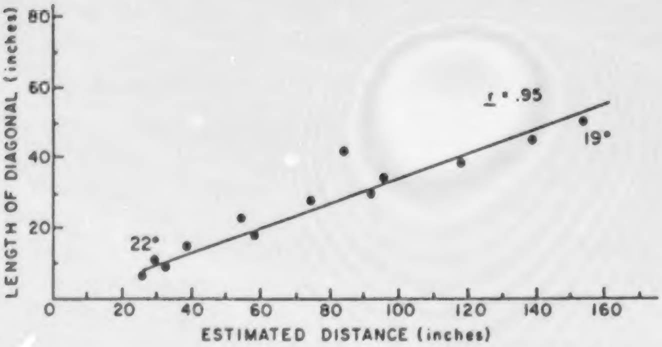


Figure 24. The average distance at which imaged objects seemed to overflow when subjects imaged line drawings of animals. (From Kosslyn, 1980, p. 78.)

(4) Images are constructed over a period of time on a part by part basis. Evidence for this conclusion comes from the result indicating that images containing several objects take longer to create than images containing fewer objects.

In Kosslyn, then, we have a detailed model of an analogical representation and a substantial amount of evidence illustrating many important features of images. Perhaps the strongest single conclusion to be drawn is that people can create images that are surprisingly veridical and that can be processed in the way that an actual picture would be processed. Imagined objects are certainly analogs of the physical objects which they represent. As we will see later, however, the matrix representational format is probably not sufficiently general for use in many cases in which we use our imagination to solve problems. It seems likely that a richer representational format is necessary.

### *Funt*

Diagrams are often valuable aids to our reasoning. We very often find it useful to construct a diagram and reason through our diagram. Given our ability to construct relatively reliable mental images, it should not be surprising that we can solve problems by constructing "mental diagrams." Funt (1980) has developed a representational system (and a computer program called **WHISPER**) in which it is convenient to represent and to manipulate "mental diagrams" for the solution of simple problems. **WHISPER** contains four basic elements:

(1) A *high level reasoner* which guides the problem solving process and produces an answer;

(2) A *diagram* which is represented by values in a matrix similar to Kosslyn's surface representation;

(3) A *retina* which can inspect the diagram and provide the *high level reasoner* with information about a transformed diagram;

(4) A set of *re-drawing transformations* which can modify an old diagram and produce a new one in which certain objects are translated, rotated, or have undergone other similar transformation.

Figure 25A illustrates a typical problem that **WHISPER** can solve. In this case, the system is to determine the nature of the chain reaction that will occur if the system of blocks illustrated in the figure were to be constructed and released. The system proceeds by first finding the major points of instability in the system. It then finds the pivot of rotation for the most unstable object. The object is then rotated about its pivot point until either the conditions for a collision are met or until the conditions for the object falling free are met. In this case, the system detects a collision (i.e., the points of two different objects fall on top of one another). At this point new

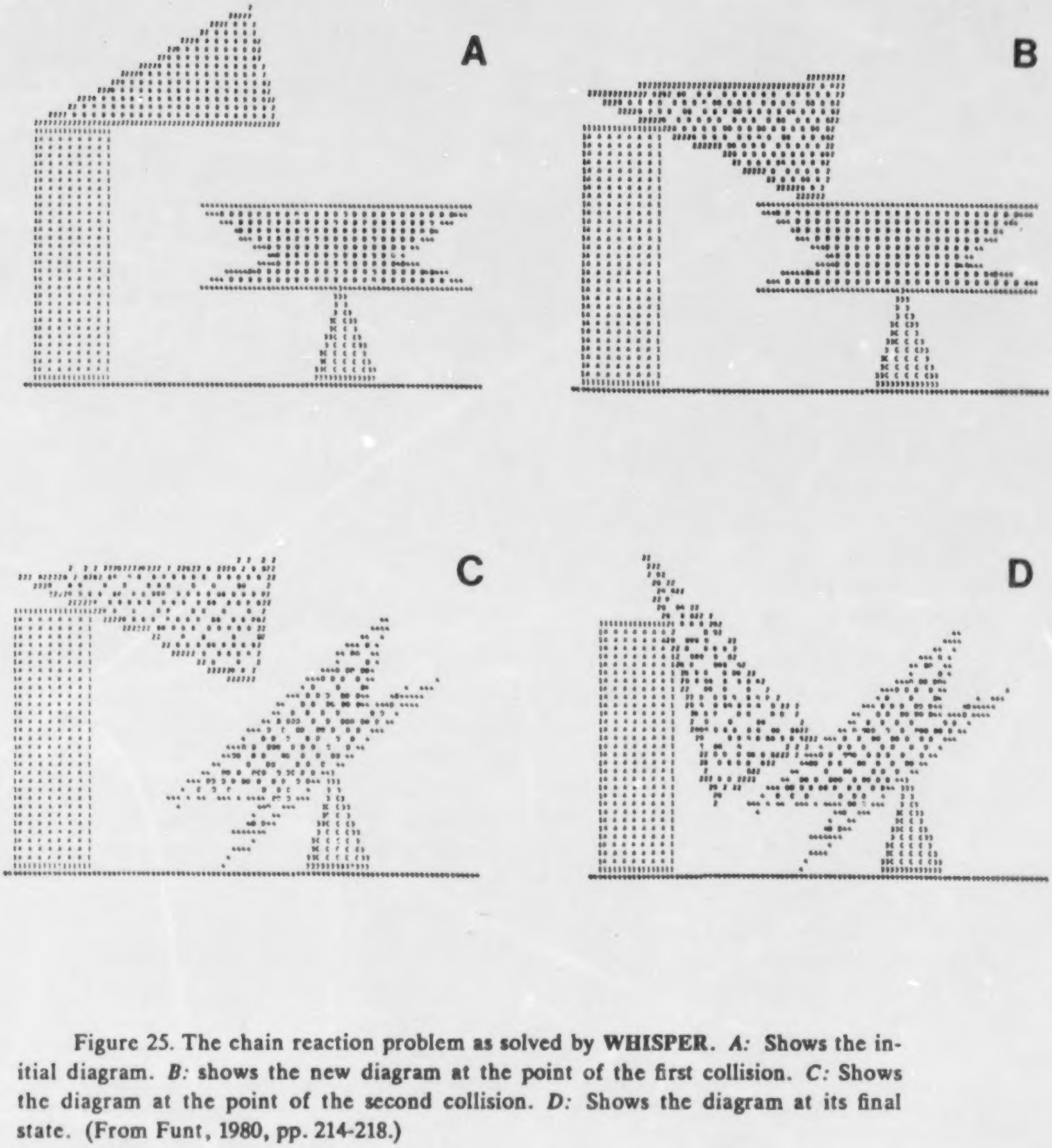


Figure 25. The chain reaction problem as solved by **WHISPER**. *A:* Shows the initial diagram. *B:* shows the new diagram at the point of the first collision. *C:* Shows the diagram at the point of the second collision. *D:* Shows the diagram at its final state. (From Funt, 1980, pp. 214-218.)

instabilities might be added, so a new evaluation is made and a new instability is chosen and followed. In the case shown in the figure, block *D* is chosen as the unstable one and the process is continued until another collision occurs (in this case, with the table). Then still another instability is chosen, finally, until no instabilities remain. Figures 25*B*, 25*C* and 25*D* show the system at the critical points at which new instabilities are sought.

It should be clear that this is a difficult problem to solve by use of equations or other similar analytic methods. In general, the "surprise collisions" can not be readily incorporated into a general solution to such a problem. Funt's method is essentially a *simulation* method in which the internal relationships of a complex system are determined through a simulation of the process. It is very often the case for complex problems that simulation is the most effective solution method. In fact, it may very well be that the essential characteristic of reasoning through imagery is that imagination is a mechanism for performing *mental simulations*. We turn now to a general discussion of the notion of a mental simulation and the more general notion of the *mental model*.

### *Mental Models and Mental Simulations*

So far, we have restricted our discussion of analogical representations to cases involving imagining. We can imagine two objects rotating and this will help us determine whether the objects are congruent. We can imagine a paper cutout being folded into a cube and answer questions about which sides fit together (Shepard & Feng, 1972). We can imagine an animal (such as a German Shepard) and use our image to verify characteristics of it (does it have pointed ears?). We can imagine diagrams similar to those used by Funt (1980) and determine the outcome of a chain reaction. We can imagine a ball rolling down a "mental roller coaster" and determine where it might end up (de Kleer, 1975). We can imagine walking through our house and determining how many windows it has. We can imagine a person pole vaulting over a high bar and just barely knocking the bar off (or just barely making it). We can imagine waking up to the smell of bacon and eggs. We can imagine the "sounds" of a symphony orchestra and "hear" the friend's response to our questions.

It is clear that our ability to imagine a wide range of activities is a very useful mechanism in our ability to reason about our world. It is not so clear, however, that a "matrix" representation is a very useful representational format for most of the cases of imagining just mentioned. In particular, we believe that rather than the "mental image" we should think of the *mental model* and rather than the "mental transformation" we should think of *mental simulations*. It would seem that the human has the remarkable ability to construct a representation of an object or situation that is a kind of *model* of the object or situation, where the model is manipulable and "runnable" as a *mental simulation*. As is usual, the decision of the kind of representation most suited to these *mental models* is a notational issue. How best can we express our theories about what these mental models are like and how best can we characterize their important features?

*Mental simulation.* One of the most important phenomena that drives the study of mental models is that of *mental simulation*. This is essentially what a billiard player must do in lining up a new shot, or, for that matter, what any skilled athlete or performer must do in determining the best course of action, be it for golf, tennis, chess, or bridge. In these situations people act as if they were running a mental simulation and observing its behavior. The "chain reaction" problem of Funt (1980) illustrated in Figure 25 is a good example, both of a problem that a person might solve by "running" a mental simulation and also of a representational system that solves the problem in much the spirit that we imagine a person would.

Consider how we might determine the functional properties of an object. It might be argued that an essential property of chairs is their "sit-on-able-ness." That is, among other things, for something to be a chair, it must be possible to sit on it. How do we determine whether it is possible to sit on something? Mental simulation often appears to be a method. Consider, for example, whether a salt shaker is "sit-on-able." Many people, when considering this example, report mentally simulating such an event, giggling at the expected outcome, but reaching an affirmative outcome when they mentally simulate either a six-inch tall human or a two foot tall salt shaker.

Mental simulations would appear to be useful devices for discovering factual knowledge buried in our tacit or procedural representations. Thus, for example, when asked a question such as "How many windows are there in your home?" people often report mentally simulating a walk through their house counting the windows.

An interesting example of the use of "mental simulation" in a computer system to facilitate the answering of questions is provided by the work of Brown and Burton (1975) and Brown, Burton, & deKleer (1982). In particular, *Sophie* could answer hypothetical questions about what would happen if a particular circuit component were changed or damaged. *Sophie* had two distinct knowledge representations about circuits. On the one hand, it had a traditional propositional representation about the causal relationships among the components, as well as principles of circuit design. In addition, however, *Sophie* contained a mathematical model of the circuit. Some questions were best answered by inferences in the semantic network while other questions were best answered by having the system set up the model of the circuit and "run", it using the results of the simulation to determine the answer. The *Sophie* system captures most of the important features of *mental models* and *mental simulations* and illustrates the power and utility of a system that has multiple representations of the same represented world. Even though a mathematical model may not be the best representation of the human capacity for creating mental models, the system serves as a powerful example of how one might combine multiple representations, including one that could be executed to deteremine the results.

*The essential features of mental models.* A detailed description of the state of the art on work in mental models is presented in Gentner and Stevens (1983). Although the work reported in this book is just the beginning of the field and the approach -- consider it a report of work in progress -- we believe that it is an important beginning for two reasons: one, as a practical aid in the design of applied systems that must reason about complex physical systems; and two, in providing a considerably richer framework than now exists for the study of mental imagery and mental transformations. These new approaches allow us to examine "images" by means of methods that do not view them as purely two dimensional visual phenomena in which a "quasi-pictorial" representation seems appropriate, but rather as part of a much broader and more important human capacity. In general, the studies of mental models reveal a number of features that seem to characteize the approach (see Gentner & Stevens, 1983 for expansion of these ideas):

- *Data and process are closely bound*. Procedural information plays a critical role in mental models, although as the work on the *Sophie* system shows, there may be both procedural and propositional representations intermixed. However, much of the power of mental models comes from their ability to simulate the represented world (by "running" the model), with the results available only by inspecting the outcome of that simulation.

- *Mentals models are likely to use qualitative reasoning*. A person's ability to reason often seems quite good *qualitatively*, but when the answers depend on a *quantitative* relation, our abilities deteriorate (in the absence of external aids). (See Forbus, 1983.)

- *Mental models usually involve causal reasoning*. Mental models are often causal models. That is, they are models which embody the causal features of the domain which the model. Thus, for example, in solving physics problems, experts often develop mental models of the physical systems discussed in the problems. These systems are abstract (in that they contain frictionless planes and other similar idealized objects), they embody the causal laws of physics in a qualitative fashion, and they can be "run" to make predictions.

- *Mental models contain a strong experiential component*. Introspection reveals that mental models contain a strong experiential component. This is why the phenomenology of imagery is also the phenomenology of mental models. It is, of course, not clear how much one should rely on introspective evidence, but it is also clear that one should not ignore it. It should be noted that the experiential component need not be visual, and if it is visual, it need not be (and probably isn't) merely two dimensional. Our imagination and mental transformations appear to contain visual, auditory, kinesthetic, and emotive components, in addition to the more abstract components necessary for the kinds of causal reasoning processes that seem to be such a fundamental part of mental simulations.

### *Propositional and Analogical Representation*

Much has been made of the supposed fundamental differences between analogical and propositional systems of representation. It is our belief that these differences are highly overstated and overemphasized. There are indeed different methods of representation, each with its own virtues and deficits, each good for a particular set of circumstances. Clearly, however, the notion of analogical representation conjures up a particular form of representation. Let us examine these aspects of representation so that we might understand how they fit into the entire spectrum of representational systems.

What does it mean for a representation to be "analogical"? In one sense, the question is meaningless, for the whole point of any representational system is that the represent*ing* world be similar or analogous to the represent*ed* world. Perhaps the best way to examine this issue is to examine the major points made in two prescient analyses of representational systems: the point made by Bobrow (1975) that there are numerous, separable *dimensions of representation* and the distinction raised by Palmer between *intrinsic* and *extrinsic* aspects of representation.

Representation is (purely) intrinsic whenever a representing relation has the same inherent constraints as its represented relation. That is, the logical structure required of the representing relation is intrinsic to the relation itself rather than imposed from outside. Representation is purely arbitrary whenever the inherent structure of a representing relation is totally arbitrary and that of its represented relation is not. Whatever structure the representing relation has, then, is imposed on it by the relation it represents. It is typical of so-called analogical representational systems that the crucial relations of the system tend to be intrinsic in the representational format. It is typical of propositional representations that the inherent characteristics of the representing relations are not characteristics of the objects being represented and thus must be added to the representation as additional, extrinsic, constraints. It should be emphasized, however, that whether a set of constraints is intrinsic or extrinsic makes no difference in the operation of the representational system. The essential feature is that representational systems have the power to express those relationships of the represented world that are being represented.

As we have already seen, the critical thing about a representation is that it maps some selected aspects of the represented world into a representing world. There are two keys to understanding the differences among representations:

1. The selection of *which* dimensions of the represented world are to be captured within the representing world;

2. The determination of *how* the selected dimensions shall be represented.

These two aspects of the decision -- the "which" and the "how" -- then govern the properties of the representational system. Note that even in the mapping of a single represented world, the questions might have to be answered several times. For each dimension of the represented world that is selected, there could very well be a different determination of how that dimension is to be represented. In some cases, the very choice of a dimension tightly constrains the set of possible ways to do the representation. In other cases, having made the one decision, there are a number of possibilities remaining for the second.

Suppose we wished to represent the star above the plus of Clark and Chase (1972), the figures that so perplexed our undergraduate students (Figure 26*A*). If we wished to represent all the spatial details of the figure, then an appropriate representational scheme might be to map spatial dimensions in the represented world into spatial dimensions in the representing world. In this case, we might set up an array of elements, letting each element in the representing world take on a value of 1 wherever the corresponding spatial location in the representing world had a light intensity less than some critical value, and being 0 otherwise: the result is shown in Figure 26*B*. For many people, this result captures the essence of an analogical representation, for the representing world looks like an image of the represented world (and this is basically the representational format used by Kosslyn, 1980, and Funt, 1980) Presumably, this representation would have satisfied our students. However, looks are not important; what matters is what can be done with the representation.

Suppose we wished to judge the relative areas of the two figures, or compare the lengths of the vertical heights, or horizontal widths, or diagonal lengths? This representation, a spatial matrix, would indeed be appropriate, for having mapped spatial attributes into spatial attributes, the relative lengths of the various dimensions are automatically (intrinsically) captured by the representation. [7] Suppose we wanted to answer Clark and Chase's question? Is the PLUS above the STAR? To do this, we would have to examine the representation, determine which set of darkened squares corresponds to the plus, which to the star, which direction corresponds to up, and make a judgement. The representation is of no particular help. That is, it is no easier to make this judgement from the representing world than from the original, represented world. Once having made that judgement, how might we record the resulting fact, namely that the star is above the plus? Well, such a fact is a proposition about the represented world, and an appropriate representation for it would be a proposition something of the form:

ABOVE (PLUS, STAR).

Note that with this propositional representation, if asked the question a second time, it would indeed help us get to the answer. However, if asked to judge the relative dimensions of the two figures, the proposition would be of no use whatsoever. Different representations have different virtues and should be used for different purposes. In general, a representation is best for purposes in which the information desired in captured in its intrinsic properties.

---

[7] There are still some assumptions that must be met. Thus, we have depicted the different elements in the representing world adjacent to one another, with the coordinate systems parallel, linearly related, and with the same scaling factor. In other situations, it might be advisable to chose otherwise, in which case the intrinsic relations that we have just relied upon for the various comparisons among dimensions might not still hold.

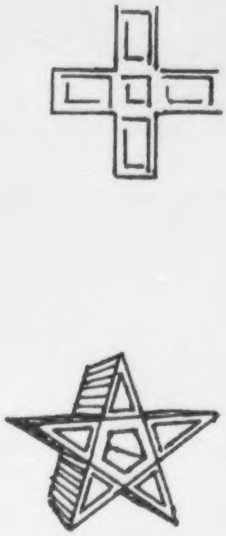


Figure 26. *A* shows an elaboration of the illustration used in the experiments of Clark and Chase (1972) in which subjects were asked to answer TRUE or FALSE to the question of whether or not the figure shows a PLUS above a STAR. In the original experiment, the stars and pluses were simple line drawings. This figure shows much more elaborate detail, intended to make the point that the characteristics of a representation are determined to a large extent by the selection of which aspects of the represented world are selected to be represented within the representing world. *B* shows a possible representation of *A*, in which spatial dimensions of A have been mapped into spatial dimensions of *B*.

Note that we represented intensity in the original world by 1's and in the depicting world by 0's. That is obviously an arbitrary, discrete representation for what could be a rich, continually varying dimension. The fact that we chose to map spatial properties into spatial representations leaves completely open the issue of how to map other dimensions, such as intensity, color, weight, odor, monetary value, etc. Again, for the purposes of this particular set of tasks, it was sufficient to represent intensity in this binary-valued, discrete fashion. Indeed, it is superior, for it means that subtle differences in intensity do not confuse our comparisons. For other purposes, such a representational choice might not be adequate.

*Analogical does not mean continuous*. One common misconception of the meaning of "analog representation" is that it is continuous whereas propositional representation is digital, or discrete. [8] This can't really be true, for although the matrix representation of Figure 26*B* would be classified as an "image" or analog representation, it clearly is composed of finite, discrete cells. Still, the notion of continuity persists, perhaps hedged with the realization that there may be a discrete cellular representation, but it is still analogical if the cells are of fine enough grain. It is easy to see where this belief comes from, for this distinction does characterize many existing systems. But the distinction is a result of the choice of dimensions from the represented world that are to be represented, not from any inherent property of the representational system itself. If we map spatial information into spatial form, then we are apt to use a continuous method of representation. If we map number of objects into either the number system or by a one-to-one map of object to representational symbol, then the most reasonable analogical representation is discrete, either the non-negative integers or finite symbols. That is, if the dimension in the represented world is continuous, then it makes sense for the representing world to be continuous. If the represented dimension is discrete -- or if the continuity of the dimension is of no particular interest -- then the best analog in the representing world would be a finite representational format. Whether or not we wish to characterize the representation as analogous depends upon how well we have captured the critical features of the represented world.

A discrete representation of a continuous dimension may still be characterized as analogical. Take the mental rotation phenomenon of two-dimensional figures as an example. First, we separate consideration of the representation of the figures to be rotated from the representation of the rotation: either one may be analogical or propositional, regardless of the other. Consider the four possibilities this gives rise to. If the figure is propositionally represented [by statements of the form ONTOPOF(cube1,

---

8. Continuity is really being confused with density here. What people often mean is that an analogical representation is dense. That is, if we represent an image of the world by means of a grid of points, then the image has the same resolution of detail as the real world: if we take any two points, no matter how close together, then there is still some other point between them. Interestingly enough, even the real world does not have this characteristic, not if we are to believe modern physics. (But of course, theories of physics are not the real world: they are simply representations of the world, but we digress.)

cube2)] [9] angular position could be represented either by discrete position [POSITION-OF(main-axis, horizontal)], or by continuous position [POSITION-OF(main-axis, 30.267 ... °)], the difference being whether the position is selected from a finite set of descriptions (such as the integers) or from the real numbers. (Levin, 1973, described how this form of representation might work for mental rotation.) If the figure is analogically represented, perhaps as in the spatial matrix form of Figure 26*B*), we still need to determine how to represent the rotation. It is easy to see how we might represent rotation in non-analogical form: we simply jump from the current position to the new position, traversing few or none of the intermediate states. If there is a matrix representation, it is not simple to actually do the rotation: the contents of each cell of the matrix would have to be moved to an appropriate new cell, and the algorithm that might accomplish this in a continuous way is not at all obvious. Yes, one could do the appropriate matrix multiplication, but then, why not just compute the desired end point -- there is no need to actually rotate the representation. Moreover, if the representation is a matrix of this form, continuity is not possible in principle, for the same angular rotation covers different numbers of matrix cells at the periphery of the figure than near the center: at some point, intervening cells must either be repeated or skipped. If we try angular rotation on a cartesian grid, the grain size problem is a fundamental limitation. A solution to this problem has been proposed by Funt (1983) who proposed using a spherical coordinate system for the representation. Funt shows that continuous rotation can be performed if a large number of processing mechanisms are packed into a spherical array, each processor communicating only with its neighbors, each containing the relevant segments of the represented figure. To perform rotation, each processor passes the relevant segments to the appropriate neighboring processor. This is true rotation, for the representation truly "rotates" through the spherical array. Note, however, that because the number of processors is finite, the rotation still takes place in discrete steps.

Consideration of what it means for the representation of rotation to be analogous to physical rotation makes it clear that the critical feature is whether or not the rotation passes through intermediate values. Indeed, this is why Shepard and Cooper (1982) place so much stress on the experimental demonstration that their experimental subjects did appear to rotate the test figures through the intermediate states. Their experimental findings allow us to conclude that people do represent rotation in a manner analogous to physical rotation. We can make this statement with confidence, regardless of whether human rotation actually is smooth and continuous, or whether it might be by discrete rotational jumps, perhaps -- as has been suggested by Just and Carpenter (1976) -- rotating in steps of 50°. As Shepard and Cooper (1982, p. 175) put it: "Just and Carpenter (1976) acknowledge that their model of mental rotation fulfills our criterion for an analog process in that during rotation of, for example, 150°, the internal process passes through intermediate stages corresponding to intermediate external orientations of 50° and 100°." The point is that we can separate the determination of something being continuous from the determination of it being analogical.

---

[9] Presumably the representation would be based upon the relationships of the component parts to some canonical position determined by the axes and centroids of the figures -- an aspect that is critical for all the representational forms.

## PROCEDURAL REPRESENTATIONS

There is a classic distinction in representational systems between knowledge about something (called *knowledge of*, or *declarative* knowledge) and knowledge about how to do something (*knowledge how*, or *procedural* knowledge). Some of our knowledge is declarative, in the sense of making a statement about some property of the world. Thus, a statement of the form "George Washington was the first president of the United States" is a prototypical declarative statement. Knowledge of how to kick a football is a prototypical piece of procedural knowledge. Declarative knowledge tends to be accessible; it can easily be examined and combined with other declarative statements to form an inference. Procedural knowledge tends to be inaccessible, being used to guide our actions, but oftentimes offering remarkably little access or ability to be examined. Thus,although we can pronounce a word like "serendipitous," we cannot say what movements our tongue takes during the pronunciation without actually doing the task and noting the tongue movements. We seem to have conscious access to declarative knowledge; but we do not have this access to procedural knowledge.

So far in this chapter we have only discussed declarative systems of representations, systems in which the manner by which knowledge is represented is the critical concern. Procedural representational systems comprise a contrasting class of systems where the concern is *what* they do, not how they do it. Note, however, that the discussion of procedural representation has intermixed two different, but related, concepts. One concern is with how we should represent the knowledge of how to do things: knowledge of how to perform actions upon the world, knowledge of mental strategies that allows us to perform actions upon the representational structures of mind. The other concern is why there is this apparent difference between the accessibility of declarative and procedural knowledge. The two issues need not be related, although in practice, they are. The first issue is actually concerned with the *representation of procedures*. The second issue is concerned with *procedural representation*. To understand the differences between these two concepts, we must first look at some of the properties of an information processing system.

### *The Human Information Processing System*

The human organism can be viewed from many perspectives, each offering different and valuable insights into our overall understanding. One important viewpoint is that of a symbol processing system, capable of manipulating, interpreting, and generating symbols to aid in its processing and understanding of itself, others, the the local environment, and the world. (See Newell, 1981, for a thorough treatment of the basic components of a symbol processing system.) The concept of a "symbol" is, of course, critical, although precise formal definition is difficult. We define a symbol to be an arbitrary entity that stands for or represents something else. By "entity" we mean anything that can be manipulated and examined. Thus, a symbol is a physical thing as opposed to an imaginary or hypothetical concept. In mammals, symbols are realized by neural signals: chemical or ionic and electrical potentials. Humans also use external devices as symbols, such as the symbols of writing and printing, electronic displays, or speech waves.

Note that the entity that is the signal is arbitrary. The marks on this page are symbols, but only because our culture has agreed upon how they shall be interpreted. Thus, not all the marks are symbols: some are not interpretable, and thus can be dismissed as "noise." Symbols alone do not suffice, for if they are to symbolize or stand for something, there must be an agreed upon convention between the symbol maker and the symbol user as to their interpretation. This, in turn, requires that there be some mechanism that can interpret symbols, manipulate them, and perform actions based upon them: we call this mechanism an *interpreter*.

Any information processing system can be conceptualized as containing a number of distinct components. There must be a system of sensors that are responsive to variations of energy flux in the environment (a sensory apparatus). There must be a system of effectors through which the system can affect the external environment (a motor system). There must be a way of storing information so that the past can affect the present (a memory system). There must be a set of processes that use both information that has been stored in memory and that is arriving currently via the sensors to determine what kinds of responses to generate and what aspects of the current state of the system will be preserved by the memory system (a processing mechanism and an interpreter). Overall, an information processing system must have five separately identifiable components:

- a sensory apparatus
- a motor system
- a memory
- a processing mechanisms
- an interpreter

Note that these five components need not be physically distinct. The processor, memory, and interpreter may use the same physical mechanisms. The sensory and motor apparatus may share mechanisms. The distinctions among these five are conceptual, not physical.

Our interest here is in the interpreter (and the symbol system upon which it operates). An interpreter acts as a translator, going from symbols to actions. An interpreter, therefore, must be capable of examining symbols and executing the actions that they specify. This means that the interpreter itself is composed of procedures. It can perform operations upon the symbols, including getting access to them, comparing them with others, and initiating actions that depend upon the results of the comparisons. Interpreters therefore use symbols in the declarative sense, for they must be able to examine the symbols and perform the operations that they specify.

### *The Representation of Procedures*

When we represent procedures in a form that is to be interpreted, then we are representing procedures in a declarative format. Consider the procedure for answering the question, "Can X fly?": [10]

```
Procedure: "Can X fly?"
        If there exists a relation can fly leading from X,
                then answer "Yes, X can fly" and stop.
        If there is no Y such that (X isa Y or X subset Y),
                then answer "As far as I can tell, X does not fly" and stop,
        otherwise, for each Y such that (X isa Y or X subset Y),
                do the procedure "Can Y fly?"
```

Note that this procedure can be represented in any of the propositional representational systems that we have examined, and, if the system had an appropriate interpreter, it could then be executed to produce the desired result. Moreover, it would even be possible to modify the representational structure according to the results found by the procedures. Thus, suppose that the representation were a semantic network. The appropriate way to do the modification is to change the first "answer" statement to read:

```
then answer "Yes, X can fly" and
if there exists a relation can fly leading from X,
        then stop,
otherwise, connect can fly to X and stop.
```

This method of imbedding procedures within the representation really means that the representational format for the knowledge in the representation (the data) and for the procedures (the programs) that operate upon the knowledge have the same format. This, actually, was a major insight of computer science in the 1940's: that it was possible to have information structures within the computer memory that could be interpreted as either data or program, whichever was relevant for the moment. This means that the same information structure can be viewed as either data (declarative) or program (procedural) -- and that is the key to this method of procedural representation. The power of this system comes from the fact that the interpreter can access procedural information as data, and thus describe it, alter it, and even simulate what would happen were the procedure to be invoked, actually doing the operations. Similarly, the interpreter can follow the procedure, thus doing the operations in the manner specified.

For many aspects of learning, the kind of accessibility provided by imbedding procedures within their own representational structure, accessible to an interpreter, seems critical. Indeed, this is what verbal or written instructions consist of: descriptions of procedures that are to be followed in performing the task that is

---

10. This is basic recursive procedure for following a semantic network hierarchy to answer a question about a property. Note that it is *not* a good model of human behavior: it will always take longest to answer that "X does not fly," which is not consistent with the observed data. Moreover, its representation of the property "can fly" is not consistent with modern systems. The procedure is being presented in order to demonstrate its format and how it gets interpreted.

being learned. The learner is expected to understand the instructions, to convert them into knowledge structures within the representational system, and then to follow them at the appropriate times in the performance of the task.

Modern algebraic computer languages (such as Algol, Fortran, Pascal, and Ada) do not allow for this kind of embedding, for they rigidly separate the data structures and the procedures that operate upon them. (Of course, the compilers for these languages do treat the procedural statements of the language, the programs, as data and transform them from a format readable and interpretable by humans into the machine language specification necessary for the computer hardware.) Many research languages, especially interpretive languages such as LISP, are self-embedded. In LISP, the data structures and the procedures that operate upon them are all written in LISP, save for a few basic primintives. The LISP interpreter is capable of understanding the procedural information, which is stated in the formalism of LISP. The schemes used in representational systems are closely related to the methods used within LISP.

One representational system to use this approach of self-embedding the procedures within the representational structures is the "active network structures" of the LNR research group (Norman & Rumelhart, 1975: hence the word *active* that modifies the term "network.") The definitions, although appearing as ordinary semantic networks, are actually procedures, that, when interpreted, carry out the necessary structure building and structure matching processes to check newly asserted information against the data base, fill unspecified variables from the context, and, when needed, build pieces of semantic network to represent the facts being asserted. (For a more complete discussion see Rumelhart & Levin, 1975.) Note that it is not enough to represent the sequences of arguments that are to be applied. Rather, one must eventually turn to some primitives, information about the actions themselves that cannot be represented at the same level as the rest of the representation (and must therefore be inaccessible to the interpreter). These primitives control the actual motor system (at least in a human: in a computer the equivalent would be the basic machine operations). Therefore, even in self embedded representation in which the procedural information is available for inspection, there is at least one kernel that is procedural in the second sense of the term: inaccessible to inspection, the view of procedures to which we now turn.

### *Procedural Representation*

In one important class of representational systems, data are stored in a procedural representation of the second sense: inaccessible to inspection. This form of representational system has certain efficiencies and other virtues. Suppose we wished a representational system to be able to answer queries of the form *"Do birds fly?"* In the representational systems that we have studied so far, that questions would be answered by seeking an explicit declaration of the knowledge, perhaps in the form of the predicate

$$Vx\ (\text{bird}(x) \rightarrow \text{fly}(x))$$

or the equivalent semantic network structure. In the preceding section we illustrated how one might search for such information within an interpreted, *declarative* system of representation. In a *procedural* representational system, the details of how the information was stored would not be visible. Instead, there would simply be a procedure available that would yield the appropriate response. Thus suppose that "bird" were a procedure (which could be thought of as a program) that could answer questions about itself. When the questions "Do birds fly" was asked, the procedure for "bird" would supply the answer: "yes" (or perhaps, "usually"). The rest of the system would have no access to the knowledge structures except through the outputs of procedures: the representational system is opaque in the sense that its contents are not visible.

There are a number of important distinctions between declarative and procedural systems, most dealing with problems of efficiency, of the control processes that are invoked in the use of the system, and with issues of modularity and accessibility of knowledge. For psychologists, it is these last issues that are of most concern -- modularity and accessibility. In a declarative system, the manner in which information is represented is of critical importance, and it is essential that the data structures be available for interpretation by other processes. In procedural representations, the data format is hidden away, inaccessible to procedures other than the one in which the knowledge is contained. All one knows is the output of the operations themselves. These differences have led to considerable argumentation and speculation about the most appropriate form of representation (see Hewitt, 1975 & Winograd, 1972, 1975).

Benefits of procedural representation include efficiency of operation, the ability to encode heuristics, and to readily incorporate both knowledge processing considerations within the same structure (see Winograd, 1975, for a good discussion of these issues). Thus many things we know seem difficult to describe in declarative fashion: we know them by the way in which we do the task. Good examples come from our skilled behavior, whether it be speech, motor control, or thought. Procedural representation allows one to tailor the way that knowledge is represented in the manner best suited for the particular task in which it will be needed. Knowledge in a declarative system must in general be useable for a variety of purposes, and it is not apt to be maximally efficient for any particular use. To many people, procedural representations seem appropriate for the the knowledge used in skilled human performance; declarative forms seem more appropriate for less skilled performance. The efficiency of procedural representations must be contrasted with the ease of inspection and modification (and thereby the ease of learning) of declarative representations. It is clear that the two different forms of representation each have their strengths and weaknesses, so that any sufficiently general system is apt to contain aspects of both.

One last point needs to be made. Any computational system -- and this includes the human information processing system -- consists of mechanisms that actually perform operations and symbols or information that specify the nature of those operations. In some sense, all knowledge is declarative up to the point where the final machinery that actually performs the physical actions is reached. Any information processing system can be thought of as being comprised of a number of levels: the representation of procedural information in declarative form at one level is translated by the mechanisms that serve as the interpreter into the procedural form -- which is thereby a declarative form for the next lower level. Thus, in writing a computer program in LISP, for example, the symbols that comprise the program are declarative in nature, being interpreted at what might be called level 1 by the LISP interpreter into some primitive "assembler" commands for the machine. These "assembler" commands, in turn, are treated as data by the interpreter at level 2, which translates them into machine language level commands. These commands must then be interpreted by an interpreter at level 3 into appropriate electrical signals which get sent to the processing unit of the computer. The processing unit, in turn, acts as a level 4 interpreter, matching appropriate patterns of voltage levels with its stored repertoire of actions, and translating the command signals into signals to the specific elements of the machine that are to do the tasks (and which might be considered to be a level 5 interpreter). The difference between knowledge that is declarative and and that which is procedural simply depends upon one's viewpoint.

*Psychological implications*. In computer systems, the act of "assembling" or "compiling" translates a declarative representation at one level of operation to a procedural representation at that level, thereby making the operations more efficient, and at the same time, less accessible from the original level. Probably from the day that an assembler or compiler was first invented, people have suggested that a major difference between skilled and less skilled human behavior is that knowledge in the skilled case has been compiled. This notion has not been pursued extensively in the psychological literature, probably because skills themselves have not been studied as heavily as other topics. The idea has recently surfaced again in a proposal by Anderson (1982).

In a series of studies, Cohen has shown that amnesiac patients can suffer severe impairments in their ability to learn new declarative knowledge, while retaining considerable learning capabilities of procedural skills (Cohen, 1981, 1983; Cohen & Squire, 1980; Cohen & Corkin, 1981). Thus, studies of two of the better studied (and most cleanly impaired) amnesiac patients, N. A. and H. M., show that although they have great difficulty in learning new declarative material, they seem to perform at an almost normal level with procedural material. For example, when N. A. was given the *Tower of Hanoi* puzzle to solve, [11] on successive days he would deny ever having experienced it before, he would complain that it was clearly a memory task that exceeded his abilities, and he would have to be talked into doing it. Yet his performance would be excellent, reaching perfect scores at about the same rate as unimpaired subjects, all while he would be stating that he did not remember how to do it. It must clearly be an oversimplification to say this, but the performance looks like a a perfect example for a handbook chapter on representation: the declarative knowledge is deficient but the procedural knowledge is normal. Because N. A. is only aware of his declarative knowledge, he denies being able to do the task, but because his procedural knowledge is normal, he can in fact do it. He can only demonstrate his knowledge by performing it. Normal people overcome these difficulties by having meta-knowledge of the contents and abilities of our knowledge structures. That is, we know what it is we know and do not know, and so we can answer questions about the competency levels of our procedures.

---

[11] Three pegs are placed side by side: name them A, B, and C. Five rings ordered in size are placed on A, biggest ring on the bottom. The task is to get all the rings to peg C, with the restriction that only one ring may be moved at a time, that a ring can be placed on any of the three pegs, but that a bigger ring can never be placed on top of a smaller one. (The number of rings can be varied.)

*Actor and object based systems.* One important aspect of procedures is how they are to get triggered: what makes them do their actions? There are basically two ways that have been suggested for invoking procedures. One is by direct invocation: some other procedure (or the interpreter) determines just which procedure it should call for the need at hand and causes it to be brought into action. The second is by a triggering mechanism: the procedure itself watches over an appropriate data base of information for data structures that are relevant to it; when the appropriate data structures exist, the procedure is triggered. (These two methods correspond to the two methods of procedural attachment used by KRL: servants -- the first method -- and demons -- the second method.) [12]

Hewitt (1975) developed a computational system using procedures that he has called "actors" that are triggered by appropriate data conditions and that communicate by sending one another messages. Actors are closely related to the general concept of object-oriented programming (as developed in Smalltalk: Kay, 1977; and now, most commonly found in LISP Machines as "flavors"). Object oriented programs represent an interesting class of representational structures in which procedures act as representational objects, each expert about domain. Each object has a set of allowable operations that can be requested by things outside the object, usually by sending messages to the object and getting an answering message in reply. Thus, the representation of "plus," "rocket-ship," or "Henry" would be handled by making them "objects," each of which has an internal state that only it knows about (or cares about). Thus, "plus" is an object that, when sent two numbers, responds by producing the sum of the numbers. In similar fashion "rocketship" can respond to messages about its velocity, direction, mass, destination. "Henry" can respond to questions about "spouse," "children," "parents," "occupation," "height," and so on. How the internal variables are represented are of no particular interest. To the outside user, the "meanings" of these data structures are given only by their actions.

Because objects serve both as data structures and as procedures that operate upon them, they can serve both as data (declarative structures) and as programs (procedures). Hewitt (1975) discusses the relevance of his actor system to the declarative-procedural controversy this way:

> Actors make a contribution to the 'declarative-procedure' controversy in that they subsume both the behavior of pure procedures (functions) and pure declaratives (data structures) as special cases. Discussions of the controversy that do not explicitly recognize the ability of actors to serve both

---

12. Hewitt (1975) points out that these two "different" methods of invoking procedures are really "completely equivalent". Nonetheless, the distinction is useful.

functions are doomed to sterility. (Hewitt, 1975, p. 189.)

In an actor or object oriented system, all data structures are objects, as are all procedures. To understand such a system, then, one has to know the following things (taken from Hewitt, 1975):

- What constitutes the natural choice of objects;
- The kinds of messages that the various objects can receive;
- The kinds of operations that each particular object can perform for each kind of message that it can receive.

One important innovation in object-based programming is offered in the "flavors" package, available on a number of LISP systems: inheritance of procedures. Much as we defined inheritance of properties (and default values) in propositional representational systems, one can define procedures (objects) whose basic kind of operations are inherited from its parents (procedures higher than it in the procedure network) and that get transmitted to its descendants (procedures lower than it in the network). This is quite analogous to inheritance the declarative systems we have already described, and further strengthens the close relationship between these objects and both procedural and declarative representations.

Although object oriented representations offer some important properties that might well be suggestive of human representational issues, to date, there have not been any investigations of these ideas from a psychological point of view. We thus cannot yet comment upon their strengths and weaknesses for psychological theory. However, there is much to commend them and, as we shall see in a minute, some of their properties have been incorporated into "production systems."

### *Demons and Production Systems*

*Demons*. An attractive processing strategy for modern representational systems is that conceptualized by "demons." Basically, it is if there were a group of active processing structures all sitting above a data base, looking for patterns relevant to themselves. Whenever a relevant pattern occurs, then the demon is "triggered," going into action and performing its activities. The results of those activities can then cause new data structures to appear in the data base, possibly causing other demons to be triggered. Alternatively, demons may pass messages among one another, or they may directly lead to sensory or motor activity. [13]

---

[13] It is not clear exactly when these structures first appeared. The predecessor for much of the work is the "demons" of Neisser and Selfridge's "Pandemonium" model of perception (1959: see the presentation in Lindsay & Norman, 1972). Not much actual work was done on these systems until recently, when the development of actor based

The reason that these processing structures are relevant to our discussion of representation is that they combine representational information with control structures. Norman and Bobrow (1976) suggested that these processing structures could be used to direct processing in such tasks as perceptual recognition, problem solving, and memory retrieval (Figure 27), and Rumelhart (1977) demonstrated how such combined processing/representational systems could lead to an "interactive" system for word recognition (Figure 27). These processing schemes are called "interactive" because they combine both data-driven (bottom-up) and conceptually-driven (top-down) processing with the appropriate representational systems. The representational systems that they use are not new; what is new is the combination of processing structure. Each schema detects arriving data that are relevant to it, processes them, and then communicates what it has found to other, higher level, schemata. This represents the bottom-up, or data driven processing. In addition, higher level schemata can direct queries to lower level ones, shaping the course of processing, seeking evidence that would confirm their relevance. (In the work of McClelland & Rumelhart, 1981, schemata also could inhibit their neighbors, so that positive evidence for one schema would also decrease the relevance of competing methods.)

Suppose that a group of schemata were attempting to recognize a printed word that had been presented to them: let the target word be *mate* (which has as neighbors such words as *date, fate, gate, late, rate, mite, mote, mute, made, make, male, mane, mare, maze*, all words that differ from the target by only one letter). The letter schemata for **M**, **A**, **T**, and **E** will all be active, each saying, "I have a --, in position --". Then, schemata for the possible words will be activated. Thus, the schemata for **MATE**, **MALE**, and **LATE**, might each see evidence that supports them, and therefore direct messages down to the lower order schemata: The **LATE** schema will enquire of the **L** schema whether it has evidence for an "L" in the first position, the **MALE** schema will ask of "L" whether it has evidence for an "L" in the third position, and the **MATE** schema will make similar enquiries. Data driven processing takes place when a schema observes data of relevance to itself and sends messages to others telling them what it has. Conceptually driven processing takes place when a schema seeks evidence that would confirm its own relevance.

*Production systems*. Production systems are a form of demon system in which all the communication among schemata takes place through a common data structure, usually called the *working memory* (WM). A production consists of an "if → then" or "condition → action" statement:

IF *(condition-for-triggering)* → THEN *(do-these-actions)*

If the conditions described on the left-hand side of the arrow are found in WM, then

---

systems, demons, the "blackboard" processor for speech recognition, and production systems all adapted various aspects of these fully or partially autonomous processing structures. Without following the history exactly, it is still clear that they are today an important conceptual tool, both for psychology and for computer science.

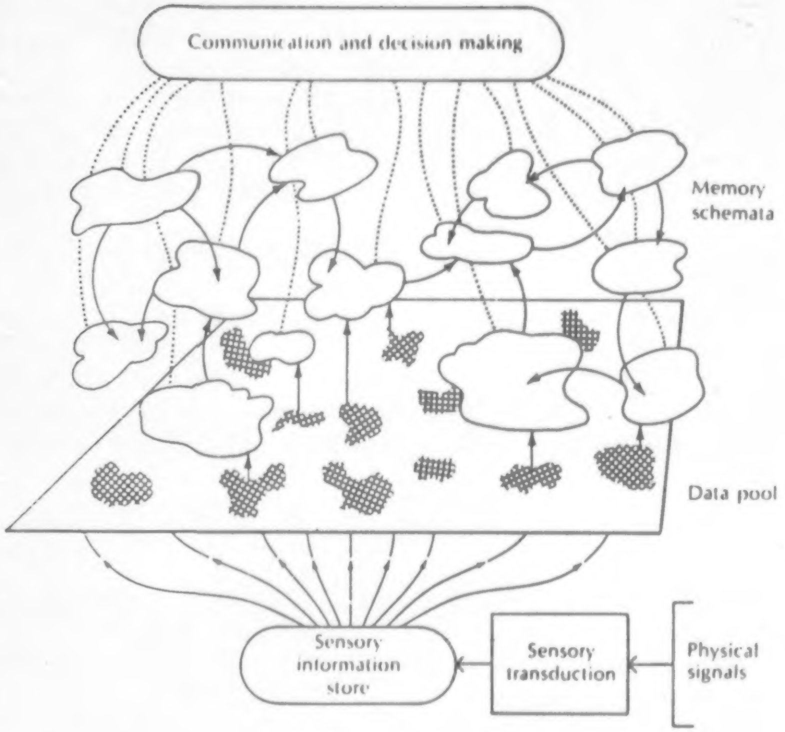

Communication and decision making
Memory schemata
Data pool
Sensory information store
Sensory transduction
Physical signals

Figure 27. The memory schemata view of the human information processing system. Incoming data and higher-order conceptual structures all operate together to activate memory schemata. Short-term memory consists of those schemata that are undergoing active processing. There is no set of sequential stages; the limits on processing capability are set by the total amount of processing resources available to the system.

do the actions described on the right hand side of the arrow. Production systems represent a form of processing called *pattern directed processing*, because the processing actions associated with a production (the procedures) are triggered into action whenever the pattern of data represented by the condition side of the production appears within WM. In general, in a production system, the actions operate upon the structures within WM, which triggers other productions to operate.

Because of the way they have been used in representational systems, production systems provide an interesting merger of active processes and control structure with representational issues. The modern use of production systems in psychology and artificial intelligence is largely due to the work of Newell (1973: the basic concept is due to Post, 1943, although it will also be recognizable as classic S-R psychology). Perhaps the easiest way to understand productions is to work through an example. Consider the productions system necessary to solve a problem in addition, such as: [14]

$$\begin{array}{r} 614 \\ 438 \\ 683 \\ \hline \end{array}$$

The productions necessary to solve any problem in addition of this type are given in Figure 28. The system works this way. First, we put the problem plus the data structure representing the goal into WM:

**goal**: do an addition problem.

This data structure matches only one production: P1. P1 is therefore activated, and it adds a new goal to WM. Note that P1 adds the new goal to the previous one. In particular, it creates a list of goals, with the new goal on top. When productions scan WM, they only see the top level goal. This type of list is called a *push-down stack*; putting a new item on the list is called "PUSHing" and taking an item off the top is called "POPping". Thus, the goal stack in WM now looks like this:

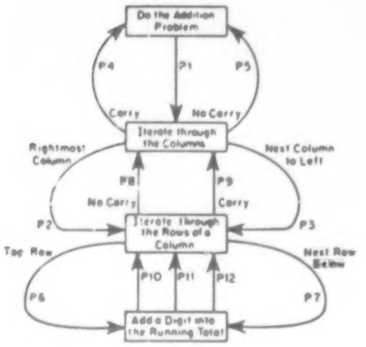
**goal**: iterate through the columns of an addition problem
**goal**: do an addition problem.

Note that only the top goal of the stack is accessible in WM. The top goal matches the condition side of production P2, and because no columns of the problem have yet been processed, P2 is invoked, PUSHing a new goal onto the stack and setting the variable "running total" to 0. Conditions are now proper for production P6 to fire, which PUSHes the goal "add the digit of the top row into the running total." Production P1, P2, and P6 have now all executed, each of them really acting to setup the structure of the problem. Working memory looks like this:

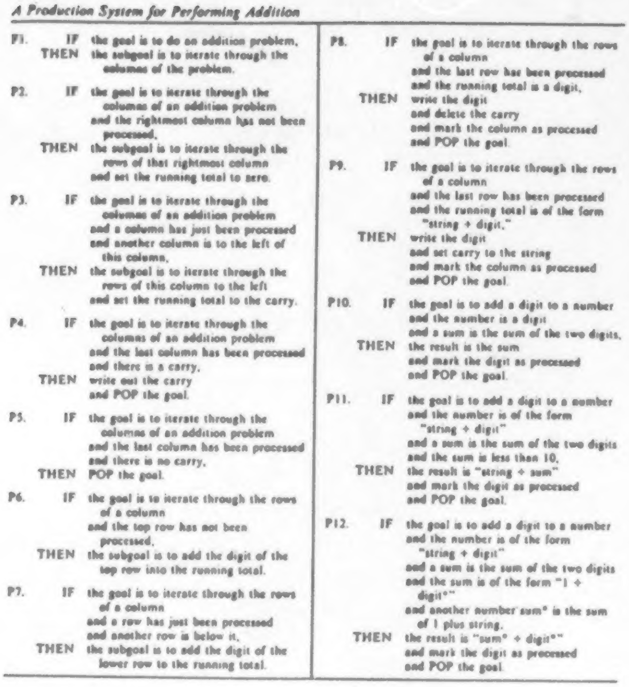
---

[14] This example is taken from Anderson (1982).

**A**



**B**



*A Production System for Performing Addition*

| | | |
|---|---|---|
| P1. | IF | the goal is to do an addition problem, |
| | THEN | the subgoal is to iterate through the columns of the problem. |
| P2. | IF | the goal is to iterate through the columns of an addition problem<br>and the rightmost column has not been processed, |
| | THEN | the subgoal is to iterate through the rows of that rightmost column<br>and set the running total to zero. |
| P3. | IF | the goal is to iterate through the columns of an addition problem<br>and a column has just been processed<br>and another column is to the left of this column, |
| | THEN | the subgoal is to iterate through the rows of this column to the left<br>and set the running total to the carry. |
| P4. | IF | the goal is to iterate through the columns of an addition problem<br>and the last column has been processed<br>and there is a carry, |
| | THEN | write out the carry<br>and POP the goal. |
| P5. | IF | the goal is to iterate through the columns of an addition problem<br>and the last column has been processed<br>and there is no carry, |
| | THEN | POP the goal. |
| P6. | IF | the goal is to iterate through the rows of a column<br>and the top row has not been processed, |
| | THEN | the subgoal is to add the digit of the top row into the running total. |
| P7. | IF | the goal is to iterate through the rows of a column<br>and a row has just been processed<br>and another row is below it, |
| | THEN | the subgoal is to add the digit of the lower row to the running total. |
| P8. | IF | the goal is to iterate through the rows of a column<br>and the last row has been processed<br>and the running total is a digit, |
| | THEN | write the digit<br>and delete the carry<br>and mark the column as processed<br>and POP the goal. |
| P9. | IF | the goal is to iterate through the rows of a column<br>and the last row has been processed<br>and the running total is of the form "string + digit," |
| | THEN | write the digit<br>and set carry to the string<br>and mark the column as processed<br>and POP the goal. |
| P10. | IF | the goal is to add a digit to a number<br>and the number is a digit<br>and a sum is the sum of the two digits, |
| | THEN | the result is the sum<br>and mark the digit as processed<br>and POP the goal. |
| P11. | IF | the goal is to add a digit to a number<br>and the number is of the form "string + digit"<br>and a sum is the sum of the two digits<br>and the sum is less than 10, |
| | THEN | the result is "string + sum"<br>and mark the digit as processed<br>and POP the goal. |
| P12. | IF | the goal is to add a digit to a number<br>and the number is of the form "string + digit"<br>and a sum is the sum of the two digits<br>and the sum is of the form "1 + digit*"<br>and another number sum* is the sum of 1 plus string, |
| | THEN | the result is "sum* + digit*"<br>and mark the digit as processed<br>and POP the goal. |

Figure 28. A production system for performing addition, consisting of 12 productions. Part *A* represents the flow of control of the productions. The boxes correspond to goal states and the arrows correspond to the productions that change these states. Control starts with the top goal. Part *B* shows the structure of the 12 productions. (From Anderson, 1982, pp. 370-371.)

**goal**: add the digit of the top row into the running total.
**goal**: iterate through the rows of the rightmost column
**goal**: iterate through the columns of an addition problem
**goal**: do an addition problem.
running total = 0

Finally, the system now does something with the problem, for the top level goal matches the condition of production P10, which not only does an addition, but for the first time, POPs the goal stack, thus removing a goal. Working memory now looks like this:

**goal**: iterate through the rows of the rightmost column
**goal**: iterate through the columns of an addition problem
**goal**: do an addition problem.
running total = 4
marked as processed: "4"

The operations continue, with productions P7, P10, P7, P11, and P9 operating in that order to complete the processing of the rightmost column, leaving the working memory in the state:

**goal**: iterate through the columns of an addition problem
**goal**: do an addition problem.
running total = "1" + 5
marked as processed: "4" "8" "3" "rightmost column"
carry = 1

Moreover, P9 puts out the partial answer: "5". The process continues until the problem is completed.

One important property of production systems is *modularity*. That is, because each production is a self contained entity, it is possible to add or subtract productions at will, without worrying about the structure of the system. As a result, new learning is readily incorporated into the system, at least in principle; as new productions are learned, they can simply be added to the existing base of productions. In practice, however, such additions are not so straightforward, and as the system gets too large, strange behavior can result from too many new additions. It seems clear that a good theory of learning is going to be required before production systems (or any other formalism) will be able to meet their apparent promise.

Production systems are destined to play an increasingly important role in the development of psychological theory, for they combine a formal processing structure of the sort that is consistent with psychological theory, plus ready implementation via a number of readily available computer programs.[15] Production systems have now been widely used in a variety of tasks, both within psychology and artificial intelligence. They form the basis for much work in artificial intelligence on expert systems, and they play a major role in such psychological work as Anderson's (1976) ACT system. A good review of production systems can be found in Waterman and Hayes-Roth (1978) and in the volumes of the *Handbook of Artificial Intelligence* (Barr & Feigenbaum, 1981, 1982; Cohen & Feigenbaum, 1982).

*Expert systems*. Determining people's knowledge structures is an old, classical problem, the basis for Freud's work, and a major aspect of clinical practice. Recently, a new application has required extensive analysis of the knowledge of experts. This is the development of *Expert Systems*, artificial intelligence systems that are capable of making progress on such tasks as medical diagnoses, geological prospecting, symbolic manipulation of equations. Many expert systems base their operation around production systems. The basic operation is to set up a basic production system architecture with sufficient power to do problem solving deduction. The hope is that by querying human experts, one can discover the rules that they follow in solving their problems, translating statements of the form:

"Whenever I see this situation, then I know that I should do . . ."

into productions of the form:

Condition → Action

The systems themselves operate by traditional production system methods, either working forwards by what is called "forward chaining" (working from what has been given, seeing what productions can be applied, then seeing what the result of performing those productions leads to, until the goal has been reached) or working backwards by "backward chaining" (starting from the goal, asking what is needed to accomplish it, using that as the new goal, and so on, until the original starting point is reached). Determining the appropriate knowledge structures to put into the system is an art, requiring skillful questioning of cooperative experts. In general, one asks experts how they solve a problem, records all that has been said, transforms the statements into productions, and then tries it out. It usually fails, because the statements of what have been encoded are incomplete and, sometimes, erroneous. At that point the expert is brought back and shown the problems. Usually the cooperative expert further expands upon the process, showing how the original statements must be qualified further and how other statements must be added. (The uncooperative expert walks out, thinking the whole exercise is a waste of time.) With each iteration, new productions are made up and added to the system, the system is tested, and the experts brought back in. The modularity principle of production systems is essential here. In the end, the systems are reasonably successful at their task, but because of the way in which it is done, it is not clear that this can really be called an exercise in showing the structure of human expert knowledge on a topic. For example, the expert has also learned a lot in the process of making the knowledge explicit. [16]

---

[15] The cost of the computers required to implement such systems is rapidly dropping; home computers will soon have this capability.

*Problems with production systems as models of human processing*. Not everyone is happy with production systems, however. Their architecture is somewhat arbitrary, and although it is claimed to match that of human processing, most of the structure had to be created in advance of good psychological theory and evidence. Working memory may correspond to human short-term memory, but the size of working memory needed to get production systems to work correctly far exceeds even the largest estimate for human short-term memory. The handling of variables seems arbitrary; we do not yet know how human processing structures manage this feat. The structure of productions is homogeneous, and does not yet match the power of the other forms of representational systems that we have studied. There are oftentimes conflicts when a number of productions simultaneously match the information within working memory, and special rules must be developed to handle these issues. And finally, the productions sometimes take on strange and arbitrary qualities, as in the first few productions of our addition example which seemed to accomplish nothing except set the stage for later ones. Not all these objections are fundamental. Most will be overcome as production systems are integrated within other forms of representational systems (for a production is really much like a "demon" of the object-based programming that we discussed earlier). Moreover, some of the problems of productions may actually be virtues; the conflicts that arise when several productions simultaneously match the conditions in working memory may be similar to conflicts that are observable in human behavior; again, see Anderson, 1982 for a treatment of some of these issues.

---

[16] References on this topic are scattered about, mostly in Technical Reports, and so the best place to start a search would be in the *Handbook of Artificial Intelligence*, Vol. 2 (Barr & Feigenbaum, 1982) and in the journal *Artificial Intelligence*.

## SUPERPOSITIONAL MEMORIES

### *Local and Superpositional Memory Systems*

One fundamental question that has major implications for theories of representation is "How is knowledge stored in memory?" Most views of memory either explicitly or implicitly assume a localized memory storage system. That is, they assume that different memories are stored in different places. Nearly all information processing systems that we understand very well have been constructed with localized memories, and it is quite plausible to assume that human memories are organized along similar lines. Thus, knowledge could be represented in the brain by local changes to individual neurons or groups of neurons. There is another possibility, however. It is possible that a given memory is distributed over many memory storage elements so that each storage element contains information from many different memories *superimposed* upon one another. This is a *distributed* or *superpositional* memory and it contrasts with *localized* or *place* storage systems. Thus, knowledge could be distributed in millions of neuronal structures throughout the brain with different data structures stored in the *same* brain structures.

Superpositional memory systems have quite different basic characteristics. In this system, different memories are not stored in separate places Rather, they are placed on top of one another, "superimposed," if you will. These systems of memory storage and retrieval offer very different solutions to some of the major issues of memory and representation. Consider the properties of the two memory systems. In *localized* memory systems:

- Different memories occupy different brain structures.
- There is a unique path or "address" that specifies how to retrieve the contents of any particular memory structure. Retrieving information, in part, consists of recovering this path information and then applying it.
- Different memory structures are stored quite independently of one another. Therefore, the physical integrity of the information within memory is not affected by what else is in memory. Of course, memory structures refer to one another by means of pointers or associations, and so they affect one another through this route. In addition, recovery of the appropriate path to a particular memory structure is made more difficult when there are many related items within the memory. But the physical integrity of the memory structures are independent of one another.

In *superpositional* memory systems:

- Different memory structures are superimposed upon one another.

- The memory structures are distributed: that is, any given memory structure must be represented across a large number of storage elements (in place memories, this is possible, but not required).
- Superpositional memories are very robust, resistant to damage of part of their memory structures. This follows from the distributed property of these memories.
- Information within the memory system is directly affected by other material. That is, in a superpositional memory system, one cannot guarantee error-free retrieval of information because of the lack of independence of storage of different items.
- Retrieving information from a superpositional memory is like detecting a signal in noise. The particular item desired is the signal, and the noise is contributed by all the other memory structures that have been superimposed on the desired one. Sometimes the signal-to-noise ratio will be high, sometimes it will be very low, hampering the retrieval efforts.
- When a known signal is presented, the system responds by amplifying the signal.
- When an unknown signal is presented, the system responds by damping the signal.
- When part of a known signal is presented, the system responds by filling in the missing parts of the signal.
- When a signal similar to a known signal is presented, the system responds by distorting the presented signal toward the known signal.
- When a number of similar signals have been stored, the system will respond strongly to the central tendency of those signals -- whether or not the signal corresponding to the central tendency has been presented.

For the most part, our ways of thinking about memory have been conditioned by our use of the local metaphor. Our language is permeated by the local view of memory. We talk about "memory search," which suggests that the memories are *someplace*, if only we could find them. We talk about *memories* as if they were *things*, suggesting a localist view of memory. For the most part, we simply adopt this view without thought. It is useful, therefore, to consider the alternative and to show how this alternative can carry out the essential tasks of a memory system.

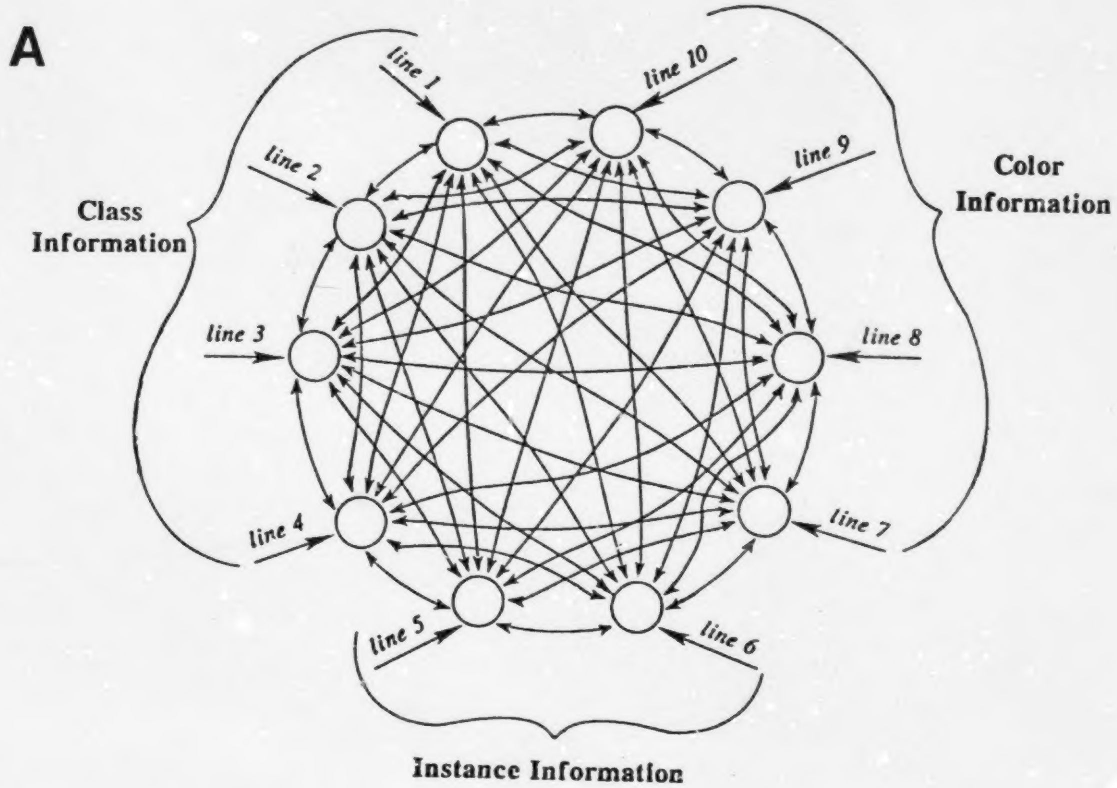
### *Associative Memories*

One major form of superpositional memory structure, called an *associative memory*, has been summarized in the book edited by Hinton and Anderson (1981). The studies reported in this book focus on the ways in which a superpositional memory might actually be realized within the brain, and so the memory structures that were examined tended to consist of a large set of simple, homogeneous, neuron-like units. A "memory," in these systems, consists of a pattern of activation across the entire set of units. Knowledge is stored in the pattern of interconnections among the units. Whenever new information is encoded in the system, those links between units whose activity patterns were similar are made larger, those links between units whose activity patterns were different are reduced in strength. Whenever the links between two units have a positive strength, we say that the two units *excite* one another. Whenever the link has a negative strength, we say that the two units *inhibit* one another. In an associative memory system, knowledge is both distributed and superimposed (additive). To say that knowledge is distributed is to say that a given concept is represented by a pattern of activity distributed over a large number of units. To say that knowledge is superimposed or additive is to say that a given unit participates in the representation of many different knowledge structures. In the simplest cases, all units are involved in the representation of all knowledge.

Perhaps the simplest way to explain these superpositional memories is by example. Figure 29*A* shows a simple ten unit associative memory system. Each unit in the system is connected to an input line and also to each other unit in the system. It is useful to imagine that each input line corresponds to a feature, perhaps the semantic features that we discussed earlier. Input lines one through four represent the category of thing being represented. Lines five and six indicate the particular class member, and seven through ten represent the color of the object being represented. The specific representations we are using for *elephants, grey things, Fido, black things, tweety bird, yellow things, dogs* and *Clyde the elephant* are illustrated in Figure 29*B*. Note, that a plus on one input line indicates that that particular feature is present, a minus indicates that it is absent and a zero indicates that the presence or absence of the feature is not specified in the input. In associative memory systems such as the one illustrated here, the input that a given unit receives is determined by the activity of units to which it is connected and by the nature of the interconnection between the units. If two units are connected by a *positive* strength, then the one unit tends to increase the activation level of the other. If two units are connected by a *negative* strength, then activation in one unit tends to *decrease* the activation of the other. Each unit responds in proportion to its total inputs and is assumed to affect other units at a rate determined by their "strength" of association. When a particular input is presented to the system it causes each unit of the system to achieve an activation level that depends upon both the input signal and the interconnections among the units. In a system with N units, the activity state of the system can be characterized as a vector of length N in which the value of each element of the vector represents the activity of the corresponding unit of the system. The pattern of interconnections of such a system can be represented by an NxN matrix, in which the *i-j*th cell of the matrix represents the degree to which unit *i* excites or inhibits unit *j*.

Information is retrieved from an associative memory in essentially two ways:

(1) A weak pattern may be presented to the system and the system allowed to respond. If the pattern had been stored in the system, then it will amplify the pattern and the final state of activation of the system will look just like


A
line 1
line 2
line 3
line 4
line 5
line 6
line 7
line 8
line 9
line 10
Class Information
Color Information
Instance Information

**B**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| elephants | (+ | – | + | – | 0 | 0 | 0 | 0 | 0 | 0) |
| Fido | (+ | + | – | – | + | + | 0 | 0 | 0 | 0) |
| bird | (+ | – | – | + | + | – | 0 | 0 | 0 | 0) |
| dogs | (+ | + | – | – | 0 | 0 | 0 | 0 | 0 | 0) |
| Clyde | (+ | – | + | – | – | + | 0 | 0 | 0 | 0) |
| grey-things | (0 | 0 | 0 | 0 | 0 | 0 | + | – | + | –) |
| black-things | (0 | 0 | 0 | 0 | 0 | 0 | + | + | – | –) |
| yellow-things | (0 | 0 | 0 | 0 | 0 | 0 | + | – | – | +) |

Figure 29. *A*: An associative memory structure consisting of ten units, each connected to an input line and to each other units. *B*: The activity patterns associated with the concepts of ***elephants, grey things, Fido, black things, tweety bird, yellow things, dogs*** and ***Clyde the elephant***. Note that colors are indicated by input lines seven through ten and the kinds of things are indicated by input lines one through six. Representations of specific individuals have non-zero values on lines four and five.

the input, except each unit will be more extreme that the input pattern. If the pattern had not been stored, then the final state of the system will be weak and different from the input pattern. This is a kind of *recognition*, in which the magnitude of the response of the system can be taken as a measure of familiarity.

(2) A type of *recall* procedure can be use in which a part of the signal can be presented and the system can *reconstruct* the original pattern from the partial cue.

One might suppose that it would be difficult to set the interconnections so that they generate this kind of behavior. However, a very simple storage procedure will lead to this pattern of behavior under rather general conditions. The simplest storage procedure of this type involves the use of the so-called ***Hebbian*** learning rule:

> If two units both respond the same way (i.e., both respond positively or both respond negatively) to a given input, then the connection between the two units should be strengthened (i.e., made more positive). If two units respond differently to a given input, then the connection between the two should be weakened (i.e., made more negative).

Figure 30 shows the connectivity matrix (set of strengths) generated by storing the patterns for "elephants are grey" (+–+–00+–+–) and "Fido is black" (++—++++—). Note that the connection between unit 1 and unit 4 is negative (–2). This is because in both patterns, the first feature and the fourth feature have opposite polarity. The connection between the the second unit and the eighth unit is positive because, in both patterns, features two and eight have the same polarity (in "elephants are grey" both are negative, while in "Fido is black" both are positive).

Now, to a first order of approximation, the output of the system to a probe can be given by taking the matrix product of the vector representing the test stimulus with the connectivity matrix. Thus, when we present the pattern for "elephants are grey" we multiply the vector (1,–1,1,–1,0,0,1,–1,1,–1) by the connectivity matrix. In this case we get, (8,–8,8,–8,0,0,8,–8,8,–8) -- an amplified version of the input vector. If, on the other hand, we present a pattern that is very different from any presented we get no response. Thus, if we present "tweety bird is yellow", (1,–1,–1,1,1,–1,1,–1,–1,1) we get (0,0,0,0,0,0,0,0,0,0). Of course, this is an extreme case, because the probed item is entirely orthogonal to any presented target. If we had presented a probe more similar to one of the stored items, we would have gotten some response out of the system.

Suppose we present the partial probe "Fido is ????". In this case we expect the system to fill in the color of Fido. Thus, we present the input (1,1,–1,–1,1,1,0,0,0,0) and we get out (6,6,–6,–6,6,6,6,6,–6,–6). We see that the response of the system is somewhat less than for the intact pattern, but that the system correctly fills in the pattern (++—) for the color -- that is the color "black" for "Fido."

Suppose we probe with the pattern for "Clyde the elephant." What color would we get back? "Clyde" was never presented, but that since "Clyde" is very similar to "elephant", we would expect the system to respond rather strongly to this input. Thus, if we probe with (1,-1,1,-1,-1,1,0,0,0,0) we get (4,-4,4,-4,0,0,4,-4,4,-4) -- that is, we get back a version of the pattern "elephants are grey." Thus, we might be able to conclude that "Clyde is grey" even though we were never presented with this input.

Superpositional memory systems seem promising models of human memories, but their potential has not yet been fully explored. It is not yet clear whether such superpositional models will displace the more traditional local view of memory in our conception of how the human memory system works.

## GENERAL ISSUES IN THE STUDY OF REPRESENTATION

### *Cognition and Categorization*

Now that we have considered a range of representational formats, it is time to think of how the things that are represented might be organized within human memory. It is easy to view the organizational problems of representation in one of two ways; we have taken both views within this chapter. One view is that the world contains objects and events, and so a major representational issue becomes how each is to be represented, perhaps by determining what features and relations are attended to and encoded by the human, perhaps by determining what primitive representational elements might be involved, and in all cases, by attempting to determine which representational format might be used. Another view is that the objects and events of the world can be classified into categories, and the representations should therefore reflect these categories so that one item might be an instance of another (hence the development of the relation *isa*), one item might be a subset of another (hence the relation *subset*), and so on. But in neither view is the emphasis on the categories themselves and just how they might be represented or related to one another. Yes, the formal tools for doing the representations of relations were discussed; but not the manner in which the human relationships might actually exist. The study of categories plays an especially important role in theories of representation and, indeed, in theories of cognition. Hence the title of this section -- *Cognition and Categorization* -- borrowed from the seminal book by that title edited by Rosch and Lloyd (1978).

Categories are neither fully artificial nor fully natural. Were they artificial, then they would be arbitrary, and the shape of existing categories would reflect the perceiver's organizational processes, driven by various internal processing matters, strategies, and communication (social) considerations. Were they fully natural, then they would exist in the world, with people acting only to perceive and thereby to encode them appropriately. The view given by Rosch and Lloyd (1978), one that we support, is that categories are neither fully natural nor artificial, but that they represent an interplay among the structured nature of items and events in the world, the processing that takes place by the perceiver, and cultural and social factors that help shape and govern a person's knowledge.

*Cognitive economy and perceived world structure*. Rosch (1978) suggests that there are two basic principles that govern the formation of categories. One has to do with cognitive economy, minimizing cognitive processing (mental work) by taking advantage of structure in the world. Thus, by recognizing that living creatures that fly all have some common features (such as wings), it becomes easier to perceive, think about, and discuss these commonalities, even though they may actually look and function quite differently from one another. Compare the wings of a mosquito with those of an eagle; the task is aided considerably by the fact that both structures are classified within the same category -- wings. The second principle asserts that the world as it is perceived already comes with structure. Some of this is a result of correlations among the objects of the world: wings co-occur with feathers more than with fur. Objects that are perceived to be "sitonable" will share more things in common than an arbitrary collection of objects. Some of the perceived structure is internal, a result of the structure of the human processing system. Thus, we can only perceive certain physical inputs, and we often add structure to the perceptions. The separate bands of the rainbow that divide it up into distinct stripes are perceived, not real: the physical structure of the rainbow is of a continually varying spectrum of electromagnetic radiation, with no breaks, discontinuities or other boundaries. In a similar way, we perceive the hues of the spectrum as a "color circle" whereas in nature, it is linear.

*Shareability constraints.* To these two basic principles of Rosch, Freyd (1983) suggests we must add a third: *shareability constraints.* Freyd points out that regardless of how we might be capable of organizing things within our minds, the necessity to share these structures with other people will necessitate a common, simplifying structure. Thus, in the determination of kinship relations, the concepts of uncle, cousin, mother, son, brother, or sister are both easily represented and easily communicated; they pass the shareability test. Different cultures share different agreed upon structures, and so what is natural and easily able to be categorized for one culture may not be for another. Thus, the Lapps have a term (*akke*) that means "father's older brother or father's older male blood relative in his generation," a categorization that does not exist in our culture. Because new concepts are described to people who do not have those concepts in terms of concepts that they already know and understand, the concepts that already exist within a culture (and for which words already exist within their language) place strong constraints on what new concepts can be transmitted among members of that culture. Moreover, Freyd suggests that "the attempt to introduce a new term that *almost* neatly fits into the pre-existing structure of the semantic domain will probably result in a distorted meaning that neatly fits into the pre-existing structure."

Freyd's hypothesis provides some interesting suggestions for a theory of knowledge representation. She points out that:

> ... it might be that the structural properties of the knowledge domain came about because such structural properties provide for the most efficient sharing of concepts. That is, we cannot be sure that the regularities tell us anything about how the brain can represent things or would even "prefer" to, if it didn't have to share concepts with other brains. (Freyd, 1983)

These three basic principles of categorization, then, to a large extent will control the sorts of knowledge structures people will develop. However, there are still a number of issues that need to be resolved. One interesting way to divide up the remaining issues is to examine separately what Rosch calls the *vertical dimension* of categories from the *horizontal dimension*. The vertical dimension reflects the *isa – superset* hierarchy, the reflection of what items belong to what other items. The horizontal dimension tells us how things at the same level of vertical organization vary. Thus, vertically, we might go from "rocking chair" to "chair" to "furniture" and to "household goods"; here, we are concerned with the features that these items have in common, or how one category is "included" in another. Horizontally (within the domain of furniture) we might go from "chair" to "table" to "bookcase"; here we are concerned with just how all these furniture categories differ from one another. Through a number of studies, Rosch and her colleagues have demonstrated that there are differences in the utility of the different levels of vertical structure, and that there is one level -- the *basic level* that tends to capture some important properties of representation.

*Basic level categorization*. We have yet to discuss how categories are formed and how they are represented. Let us briefly return to the formalization provided by Tversky earlier in this chapter (Tversky, 1977; Tversky & Gati, 1978). We can state the measure of similarity between two sets, A and B, by an expression of the form:

$$f(A\cap B) - [\alpha f(A-B) + \beta f(B-A)]$$

where $f(X)$ is a measure of the salience of the features in set $X$ and $\alpha$ and $\beta$ are constants. This expression states that the similarity is a function of what the two sets have in common ($f(A\cap B)$) minus the ways in which they differ (the features in $A$ but not in $B$, $A-B$, and the features in $B$, but not in $A$, $B-A$). Rosch proposes that *basic level* categories are those that *maximize* the similarity of things within the category and that have *minimized* the similarity of things between categories.

Consider the categorization of furniture. The category "furniture" is not at the basic level: things within the category (chair, table, bookcase, picture, clock) do not share many features in common. The basic level is one level down. Thus, "chair", "table", and "bookcase" are basic level items. Consider chairs: they share much in common with one another; they tend to look the same, have the same function, similar size, and so on. Moreover, chairs are quite distinct from the other members of the furniture category; chairs don't look the same or function the same as tables, pictures, or clocks. At a lower level, different categories such as "armchairs" or "rocking chairs," are quite similar: there is not much distinctiveness between categories. Thus, all rocking chairs may tend to look and act in a similar way, but they are also similar in appearance and function to armchairs, dining-room chairs, and office chairs. It is only at the basic level that we simultaneously maximize similarity within and differences between category members. Rosch argues that basic categories can be determined by examining the attributes that items have in common (or in distinction), differences and similarities in motor movements when using the items, and in their shapes.

Rosch suggests that basic level categories play a major role in processing and in the organization of knowledge. One role they play is that of prototypes, helping to classify new experiences, and then helping to form a new encoding. Rosch argues that the basic level has implications for at least four different things:

Images. The basic level is the highest level for which a person can form an image of the class. That is, it is possible to form an image of your favorite living-room chair, or of living-room chairs in general, or even of chairs in general, but it is not possible to form an image of one piece of furniture that is not also a basic level (or lower) exemplar of furniture. Basic and lower level categories can have images that represent the entire class: higher levels cannot.

Perception. Consider the perception of an object at a distance: small, fuzzy, not readily identifiable. Suppose the object is in the distance coming towards you, on the ground. At first it is unidentifiable, although the fact that it is visible travelling on the ground at a certain distance and speed restricts the set of possibilities. Rosch argues that the first identifiable level at which an object can be identified is the basic level ( See Smith, Balzano, & Walker, 1978).

Development. Because perception, motor movements, functions, and images all lead to the same level of categorization, Rosch argues that "basic objects should be the first categorizations of concrete objects made by children" (Rosch, 1978, p. 38).

Language. Finally, basic level items tend to have single-word names and tend to be the level at which something is described (unless there is a communicative need to be more specific or more general), so that in describing a general object, such as an animal in the park, one is apt to call it a "dog" rather than an "animal" or, more specifically, a "yellow labrador retriever." In American Sign Language (Newport & Bellugi, 1978), it is basic-level categories that are most often coded by single signs, and super- and sub-ordinate categories that are likely not to have any sign encoding.

Despite these processing implications, Rosch argues that the notion of basic level categories is most important for the culture, not necessarily so important for a particular individual's processing and representational structures. That is, individuals develop their internal representational structures as a result of the particular experiences that they have had. Basic level structures are of more importance to the culture and the language. Freyd's "shareability" notion, suggests how the transfer between the concepts acquired by an individual and the concepts held by the culture may take place.

*How are categories defined?* Recent advances in our understanding of categorization have made it clear that we cannot expect most natural categories to have clear, rigid definitions. That is, we should not expect that we can always find clear, definite rules that allow us to determine exactly what the members of any particular category are. Yes, some categories are well defined, such as the concept of a "square." In general, however, we find that category members include some clear exemplars -- things that nobody would dispute are members of the category -- and some rather marginal exemplars -- things that are greatly disputed and for which even one person may vacillate from moment to moment. Determining category membership is much like determining whether a particular sample of time should be defined as "night" or "day": we think we understand the difference and the instances are clear cut, as long as we stick to instances near mid-day or mid-night and do not have to deal with the boundaries at dusk and dawn. Matters are even less clear if we are asked to define the categories "dusk" and "dawn".

It is, of course, not a new finding that category membership can be an ill defined concept. Within philosophy, the point has long been made, Wittgenstein (1953) being perhaps the prototypical example. Given that firm boundaries cannot be established to define category membership, how then is membership to be defined? There are numerous possibilities. One point of view is that the *classical* definition should be the starting point: all instances of a concept share common properties -- call these the *defining* properties -- and category membership is simply determined by whether or not any particular instance has all of the defining properties. From this starting point, one can then argue that the concept of membership in the category should not be determined by classical logic, but rather by alternative rules. One major alternative is to use the mathematics of *fuzzy set theory* or *fuzzy logic* to define the degree of category membership of any particular instance (Zadeh, 1965; Oden, 1977). One approach is to assume that each category has some general, *prototypical* member, and category membership is determined by how well any particular instance matches the prototype. Another approach is to argue that there is neither a set of defining features nor a prototype, simply examples of category members. Overall, there are numerous approaches, and perhaps numerous solutions, but as yet, no common agreement exists on the appropriate methods for representing human categorization. (Smith & Medin, 1981 offer a good review of many of the approaches.)

*Prototypes*. Rosch (1978) and Rosch and Mervis (1975) define prototypes to be "the clearest cases of category membership defined operationally by people's judgments of goodness of membership in that category." The prototype member of a category does not really have to exist. Thus, the prototype "animal" for American university students might be a four-legged animal with fur, a tail, size somewhere between a large dog and a cow, and other features borrowed or adapted from a variety of actual animals. No single existing animal may match the prototype. Rosch believes that the prototype probably develops in much the same way as the basic level category develops: the prototype is formed so as to maximize its similarity to the other members of the category while also maximizing its difference from the prototypes of other, contrasting categories.

The notion of prototype has important implications. People do not act equally towards all members of a category. "Robin" is a more "typical" bird than are "chickens," "ducks," or "penguins." "Murder" is a "typical" crime, whereas "vagrancy" is not. People are much faster at determining category membership for typical members than for non-typical members. Rips, Shoben, and Smith (1973) found that to American college students, "mammal" and "animal" meant almost the same thing, that "typical" animals were thought of as having four legs and being warm-blooded. Not only does this make a person a non-typical animal, but insects, lizards, and other creatures are far from the central prototype of "animal." As a result, when one thinks of a category, one thinks of the things like the prototype. One is therefore apt to attribute characteristics to the entire category that actually apply only to things like the prototype. This is an obvious source of error.

Prototypes can aid in the determination of category membership. One processing rule that captures much of the flavor of prototypes is to determine the similarity of the instance that is to be judged to all possible prototypes; the prototype that is most similar to the instance determines its categorization. This is a version of the "nearest neighbor" rule; if you imagine the prototypes as points in a multi-dimensional space, where the dimensions are the possible features, then the instance to be judged can also be represented by a point, and its categorization is determined by which prototypical point is closest to it. The rule of similarity, however, is richer than the multidimensional "nearest neighbor" rule because it allows for non-dimensional considerations such as "fuzziness" or probabilistic characterization of the variables. Note too that the rule of similarity allows for the various features and aspects of similarity to be weighted differently at different types, so that depending upon the circumstances (that is, the context in which the judgement is being made), the same instance could be categorized differently.

### *Generalization*

A pervasive tendancy of human thought is to generalize, to act as if general truths exist on the basis of experiences with a limited number of examples. The tendancy is strong enough that we can believe that we have been given specific evidence for the generalization, even though we have not. Posner and Keele (1968) demonstrated that when subjects are shown dot patterns that are distorted versions of a prototype, they learn to classify them quite well, generalizing across the various presentations. More important, the subjects judged the actual prototype to be the best exemplar of the category and believed that they had been presented with it, even though they were never shown the prototype during the training trials. A similar finding has been reported for people's memory for sentences (Bransford & Franks, 1971) and for the characteristics of members of social clubs (Hayes-Roth & Hayes-Roth, 1977).

Two important by-products of generalization are *overgeneralization* and *overdiscrimination*. In overgeneralization, too many things are classified as an instance of the category; in overdiscrimination, not all members of the category are properly classified. Thus, if we were to classify all "animals that fly" as "birds" we would overgeneralize, for we would falsely include bats and flying fish. If we were to believe that "all chairs have legs," we would overdiscriminate, for we would thereby exclude chairs that hung from the ceiling, chairs on pedestals, and bean-bag chairs. Perhaps the most famous cases of overgeneralization and overdiscrimination occur in the study of the categories of developing children who have been reported to do such things as call all men "daddy" or use the term "doggie" only to refer to the family dog. In the learning of the inflections of language, we can find overgeneralization and sometimes "oscillation." Thus, the child might first learn the proper past tense of a particular verb, thereby using verbs like *give* and *gave* properly. Then the child learns that past tenses are formed by adding "ed" to the verb, leading to overgeneralization; the past tense of *give* is spoken as *gived*. Eventually, the child learns not to apply the generalization to all possible instances. The pattern of responses therefore "oscillates":

give → gave
give → gived
give → gave

In part because of the general importance of the phenomena, the issues of generalization are important testing grounds for theories. In this section we demonstrate the differences among representational theories by discussing three different ways of handling generalization.

*Generalization through the formation of generalized schemata.* Perhaps the easiest way to begin is to consider how one of the standard schema-based theories handles generalization. The basic principles are fairly straightforward and have even been incorporated into introductory textbooks (Lindsay & Norman, 1972). The essence is that concepts are generalized whenever a number of different concepts share a sufficient number of attributes. The generalization takes place by forming a new schema -- the generalized schema -- that acts as a superset of the instances to be generalized. This forms a new class of elements -- a category -- and through the principle of inheritance of properties, from then on all instances of the class inherit the appropriate generalized properties. Thus, whenever a new instance is added to the category, it automatically inherits the generalized properties by default unless specific information is available to indicate otherwise.

Note that this model can easily lead to the phenomena of overgeneralization and overdiscrimination. Thus, if the generalized schema is not sufficiently specific, it will match a large number of instances, thereby leading to the inclusion of too many things into its class (giving the wrong default values). This is overgeneralization: applying the concept to too broad a range of exemplars. If the generalized schema is too specific, having too many restrictions on what it requires of its exemplars, it will not match a sufficient number of instances, thus leading to overdiscrimination.

This model is, in many ways, the prototypical model of generalization. It is difficult to get data that would discriminate between this model and other alternatives, but because this is such a natural way to handle generalization, it is the natural starting place, the model against which all others must compete.

*Generalization without specific generalized concepts.* There is no real need to form a specific generalized schema to represent the generalization of a concept. The issue here really is the relationship between the information within memory and the information implicit within the procedures that operate upon the memory structures. Representational issues really require consideration of the doublet of representational structure and procedure; information can be traded between the explicit structure and the procedures that operate upon the structures. So it is with generalization. If the procedures contain the proper mechanisms, the generalizations can always be performed on the fly, when needed, from whatever information is already present in the data base. Thus, suppose we have four specific exemplars of something: call them **A**, **B**, **C**, and **D**. We could generalize these four exemplars by forming a specific generalized schema, **G**. But suppose, instead, that we simply keep the specific examples. Whenever we need information about things with attributes of these schemata, we could procedurally operate upon the memory structures, and compute the desired information. In this way, generalization would occur without any need for an explicit generalized schema to be formed. Moreover, the outside observer could not distinguish this schema from the one in which a particular generalized node existed.

Basically, the difference between this method of forming generalizations and the preceding method is exactly the difference between declarative and procedural representations: the difference is solely in the availability of the information and the efficiency of the operation; to the observer, the two processes cannot be distinguished.

*Superpositional models of generalization.* The difference between "place" and "superpositional" memory storage also leads to a difference in how generalization might get accomplished. Generalization falls readily out of superpositional representational models. Thus, McClelland (1981) has shown how it is possible for a superpositional model to generalize the general attributes of class members without having any explicit generalized schema. This model differs from the procedural model just discussed only in that the distinction between the memory representations and the procedures are not clearly marked, for in the superpositional model, the procedures act on the representations through activation.

McClelland's model can be considered to be a cross between the normal schema-based models and the full superpositional memory system. The most serious problems with this account involve its lack of a type-token distinction. Thus, it is difficult to prevent generalized values from being associated with instances even where it is clearly known that the normal default does not apply. McClelland examined the distribution of members in two different hypothetical social clubs (this is similar to the situation studied by Hayes-Roth & Hayes-Roth, 1977). Thus, if most members of the "jets" wear glasses, but one member (Helen) does not, it is difficult to prevent this model from asserting that even Helen wears glasses. This "overgeneralization" is actually reasonable, for we would expect people to have problems with this fact, but, of course, they would eventually be able to learn the current situation. In McClelland's model this ability requires the development of more distinguishing features so that *Helen* would be different enough from the other members of the group to stand out as a distinct individual.

## CONCLUSION

The problem of representation is one of determining a mapping between the concepts and relations of the *represented* world and the concepts and relations of the *representing* world. The problem for the psychologist, of course, is to find those representational systems that cause the behavior of our theories to correspond to the behavior of the human. In developing a theory of representation, it is important to be aware of exactly what it is that is being represented: in particular, much of cognitive psychology and artificial intelligence is concerned with attempts to represent the mental activity of the human. To quote the earlier portion of this chapter: "within the brain, there exist brain states that are the representation of the environment. The environment is the represented world, the brain states are the representing world. Our theories of representation are in actuality representations of the brain states, not representations of the world."

In many ways, the "representation problem" is, in truth, a "notation problem." That is, in establishing a representation for our theories, we wish to discover a notation:

(1) That is rich enough to represent all of the relevant data structures and processes;

(2) In which those processes which we wish to assume are natural (i.e., are easily carried out) are, in fact, easily carried out.

### *Three Major Controversies*

Traditionally, the problem of representation has had a number of different components that have led to long debate. Three major debates have arisen over the distinctions between representational formats: propositional versus analogical, continuous versus discrete, and declarative versus procedural. The position that we have taken in this chapter is that these debates do not reflect fundamental distinctions about representational systems, but rather reflect differences in the way that representational systems meet the two criteria for such systems stated above. Let us review each issue briefly.

*The propositional -- analogical controversy.* Propositional representations are ones which consist of formal "statements" that reflect the represented world, either in the form of networks, schema-based structures, or logical formulae. Analogical representations attempt to determine a "direct" mapping between the characteristic of the represented world of primary importance and the representing world. Thus, spatial or temporal properties of the represented world might be mapped onto spatial properties of the representing world, and ordered properties of the represented world are mapped onto ordered properties of the number system in the representing world. All representational systems are, of course, to some extent analogs of the represented world; after all, that is what a representation is all about -- to capture the essence of the represented world. Whatever the mapping, a key feature of representations that we are willing to call *analogical* is that if the thing being represented undergoes change or modification, then the structure in the representing world should undergo the corresponding change or modification, passing through the same intermediate states as the original. Thus, if we have a picture of a star above a cross and move the star closer to or further from the cross, an analogical representation of that movement will have to represent the same set of intermediate states as the physical movement. This could be accomplished with a representation that consisted of a manipulable "picture" of the star and cross, perhaps in a matrix or "bit map," or it could be represented by using a two-dimensional coordinate system within a set of propositions, specifying location by values on the real numbers.

A useful way to view the differences between analogical and propositional representation is to map it into the distinction raised by Palmer (1978) between information that is "intrinsic" to the representation and that which is "extrinsic." We say that a representation is an *analog* of the represented world when the relations of interest to us are "intrinsic" to the representation.

*The continuous -- discrete controversy.* Oftentimes, continuous representations are confused with analogical, and discrete with propositional representations. However, the two distinctions are actually independent of one another. What is involved here is the "grain size" or "acuity" that one wishes to have in the represented world. Thus, if the things to be represented are discrete in nature, then even the most analogical representation in the representing world is likely to be discrete. Alternatively, one might chose a continuous (real-number) representation within a propositional structure. The real point is that one is attempting to capture aspects and relations that are considered important in the represented world within the structures of the representing world, and the choice of a discrete or continuous representation simply reflects the choice of what features are important. Thus, if one represented a moving object by a matrix representation of the object, where the movement was represented by small, discrete changes in the the representing location, this would qualify as an analogical representation as long as the discrete steps within the representing movement were small relative to the step size of interest. In this case, a discrete representation of a continuous event would still be considered analogical.

*The declarative -- procedural controversy.* The difference between representations called "declarative" and representation called "procedural" really reflect differences in the *accessibility* of the information to the interpretive structures. In the the case of declarative representations, the information is represented in a format that can be examined and manipulated directly by the interpretive processes. Thus, the information is accessible for inspection, for use by multiple processes, and for that matter, for the interpreter simply to announce whether or not the information is known to be present within the representational system. In the case of procedural representations, the information is not available in a form that can be accessed by the interpreter. Rather, one must "execute" the procedure and examine the results. Information that is procedural is therefore "encapsulated" for this level of representation, not available for inspection, not easily available for multiple processes (unless their use has been explicitly provided for), and it is not possible for the interpreter to make announcements regarding the presence or absence of information that is procedurally encoded. Declarative information is "explicit" in that it is directly encoded. Procedural information is "implicit" in that the procedure has to be executed in order to get the information.

In this chapter we have argued that what is declarative and what is procedural information is context dependent. That is, any realistic information processing system has several levels of processing and interpretations, and what is procedural at one level of interpretations is most likely declarative at a different level -- indeed, at the level where some interpretive process operates upon the procedure in order to execute it. The system is eventually grounded in the primitives of the system and in actual physical actions. And at this level, all the actions of the system are "procedural."

### *Data Structure and Process*

Representational system consists of at least two parts:

- The data structures, which are stored according to some representational format;
- The processes that operate upon the data structures.

Much confusion has arisen in the comparison of representational systems because of a lack of recognition that both data and process are essential; one cannot be understood without reference to and understanding of the other. Note that the distinction between data structures and interpretive processes varies with different modes of representation. Thus, one difference between declarative and procedural representations has to do with the relative tradeoff between the division of the knowledge between the data structures and the interpretive system. In the superpositional structures, the two different aspects are merged into the same structures, so that the interpretive structures *are* the data structures. In all cases, both need to be considered in order to understand the representational system. Data structure and their interpretive processes are intrinsically intertwined; the two must be considered as an inseparable pair in determining the properties and powers of the representation.

### *Multiple Representations*

There is no single answer to the question "how is information represented in the human?"; many different representational formats might be involved within the human representational system. Thus, within the representing world, different aspects of the represented world might be represented through different representational formats. This allows each dimension to be represented by the system that maps best into the sets of operations that one wishes to perform upon them. Different representational systems have different powers, and the choice of which one is used reflects those powers.[17]

Like every other representational decision, the decision to use multiple representations of the same information has its tradeoffs. In this case, the extra powers must be traded off against the problem of coordinating the information in the separate representations, so that when a change is made, all structures are properly synchronized so as to reflect the same represented world.

*Virtual knowledge.* Procedural, declarative, analogical, propositional -- these different terms refer to different choices in the representational format, different decisions as to which information is to be represented "intrinsically" and which to be "extrinsic," which to be "explicit," and which to be "implicit." Analogical systems are those in which the mapping of the intermediate states of the representing world correspond to the intermediate states of the represented world. Procedural systems are those in which the interpretive processes have access only to the products (results) of "running" the representation.

One of the problems in attempting to assess a person's knowledge structure is that some of that knowledge may be directly represented, and some may be indirectly coded, inferred or otherwise generated at the time of test. Modern representational theory -- as represented by the discussions in this chapter -- provides a rich set of possibilities for the possessor of knowledge. The research recognizes that people have the capability of making new inferences even as they answer a query, that much of what is reported may be generated, on-line, in real-time, at the time of answering the questions put to them, using the representational properties of inheritance and logical inference, and using prototypical schemata to structure the organization of what is being generated, complete with default values. The possessor of the knowledge itself cannot distinguish between memory retrievals that are regenerated on the spot according to some generic properties and memory retrievals that are accurate reflections of the actual events. Finally, the problem of determining a person's memory structures are amplified by the fact that much knowledge may be represented procedurally, and procedural knowledge -- by definition -- is inaccessible to its possessor.

---

[17] See the discussions by R. J. Bobrow & Brown (1975). D. Bobrow (1975) emphasizes the differences among different dimensions of a representation. And note the mixed mode format that Kosslyn (1980) uses to represent mental images.

# References for

# REPRESENTATION IN MEMORY

David E. Rumelhart Donald A. Norman
Department of Psychology
and
Institute for Cognitive Science
University of California, San Diego
La Jolla, California 92093

## GENERAL REFERENCES AND SOURCES

There are a number of good general sources for more thorough treatment of the issues discussed in this chapter. We recommend two handbooks:

- *The Handbook of Learning and Cognitive Processes*, especially Volumes 4 and 6 (Estes 1976, 1978).
- *The Handbook of Artificial Intelligence*, Volumes 1, 2, and 3 (Barr & Feigenbaum, 1981; Barr & Feigenbaum, 1982; Cohen & Feigenbaum, 1982).

In addition, see the book that started much of the work on representation in memory: Tulving and Donaldson's *Organization and Memory*, (1972). Two important collections of papers are Bobrow & Collins's *Representation and Understanding* (1975), and Rosch and Lloyd's *Cognition and Categorization* (1978).

---

These are references for the chapter "Representation in Memory" for the revision of the Steven's "Handbook of Experimental Psychology": R. C. Atkinson, R. J. Herrnstein, G. Lindzey, and R. D. Luce (Eds.), *Handbook of Experimental Psychology*. Wiley: in preparation. Comments are welcomed.

## REFERENCES

Abelson, R. Concepts for representing mundane reality in plans. In D. G. Bobrow, & A. M. Collins (Eds.), *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 1975.

Anderson, J. R. *Language, memory, and thought.* Hillsdale, N. J.: Lawrence Erlbaum Associates, 1976.

Anderson, J. R., Acquisition of cognitive skill. *Psychological Review*, 1982, *89*, 369-406.

Anderson, J. R., & Bower, G. H. *Human associative memory.* Washington, D. C.: Winston, 1973.

Barr, A., & Feigenbaum, E. A. (Eds.), *The handbook of artificial intelligence, vols. 1 & 2.* Los Altos, Calif.: William Kaufmann, 1981, 1982.

Bartlett, F. C. *Remembering.* Cambridge: Cambridge University Press, 1932.

Bobrow, D. G. Dimensions of representation. In D. G. Bobrow & A. M. Collins (Eds.), *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 1975

Bobrow, R. J., & Brown, J. S. Systematic understanding: Synthesis, analysis, and contingent knowledge in specialized understanding systems. In D. G. Bobrow & A. M. Collins (Eds.), *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 1975.

Bobrow, D. G., & Collins, A. M. (Eds.). *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 1975

Bobrow, D. G., & Norman, D. A. Some principles of memory schemata. In D. G. Bobrow & A. M. Collins (Eds.), *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 1975.

Bobrow, D. G., & Winograd, T. An overview of KRL, a knowledge representation language. *Cognitive Science*, 1977, *1*,, 3 - 46.

Bower, G. H., Black, J. B., & Turner, T. J. Scripts in memory for text. *Cognitive Psychology*, 1979, *11*, 177-220.

Brachman, R. J. On the epistemological status of semantic networks. In N. V. Findler (Ed.), *Associative networks: Representation and use of knowledge by computers.* New York: Academic Press, 1979.

Bransford, J. D., & Franks, J. J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, 2, 331-350.

Brown, J. S., & Burton, R. R. Multiple representations of knowledge for tutorial reasoning. In D. G. Bobrow & A. M. Collins (Eds.), *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 1975.

Brown, J. S., Burton, R. R., & deKleer, J. Knowledge engineering and pedagogical techniques in SOPHIE I, II, and III. In D. Sleeman & J. S. Brown (Eds.), *Intelligent tutoring systems.* London: Academic Press, 1982.

Clark, H. H., & Chase, W. G. On the process of comparing sentences against pictures. *Cognitive Psychology*, 1972, *3*, 472-517.

Cohen, N. Neuropsychological evidence for distinction between procedural and declarative knowledge in human memory and amnesia. Unpublished PhD thesis. University of California, San Diego, 1981.

Cohen, N. J. Amnesia and the distinction between procedural and declarative knowledge. In N. Butters & L. R. Squire (Eds.), *The neuropsychology of memory.* New York: Guilford Press, 1983.

Cohen, N., & Corkin, S. The amnesic patient, H. M.: Learning and retention of a cognitive skill. *Society for Neuroscience Abstracts*, 1981, *7*, 235.

Cohen, P. R., & Feigenbaum, E. A. (Eds.), *The handbook of artificial intelligence, vol. 3.* Los Altos, California: William Kaufmann, Inc., 1982.

Cohen, N., & Squire, L. R. Preserved learning and retention of pattern analyzing skill in amnesia: dissociation of knowing how and knowing that. *Science*, 1980, *210*, 207-209.

Collins, A. M., & Loftus, E. F. A spreading activation theory of semantic processing. *Psychological Review*, 1975, *82*, 407-428.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, *8*, 240-247.

Collins, A. M., & Quillian, M. R. Facilitating retrieval from semantic memory: The effect of repeating part of an inference. in A. F. Sanders (Ed.), *Attention and performance III*. Amsterdam: North Holland, 1970. (Also published in *Acta Psychologica*, 1970, *33*, 304-314.)

Cooper, L. A. Demonstration of a mental analog of an external rotation. *Perception & Psychophysics*, 1976, *19*, 296-302.

Cooper, L. A., & Podgorny, P. Mental transformations and visual comparison processes. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, 2, 503-514.

deKleer, J., Qualitative and quantitative knowledge in classical mechanics, M.Sc. Thesis (1975), MIT, Cambridge, MA.

Fahlman, S. E. Representing implicit knowledge. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory*. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1981.

Feldman, J. A., & Ballard, D. H. Connectionist models and their properties. *Cognitive Science*, 1982, *6*,, 205-254.

Fillmore, C. J. The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory*. New York: Holt, Rinehart, & Winston, 1968.

Forbus, K. Qualitative reasoning about space and motion. In D. Gentner & A. L. Stevens (Eds.), *Mental models*. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1983.

Frege, G. Uber Sinn und Bedeutung. *Zeitschrift fur Philosophie und philosophiche Kritik*, 1892, *100*, (all). (Also see P. Geach & M. Black, (Eds.), *Translations from the philosophical writings of Gottlob Frege*. Oxford: Basil Blackwell, 1960.

Freyd, J. J., Shareability: The social psychology of epistemology. *Cognitive Science*, 1983, in press.

Funt, B. V., Problem solving with diagrammatic representations. *Artificial Intelligence*, 1980, *13*, 201-230.

Funt, B. V. A parallel-process model of mental rotation. *Cognitive Science*, 1983, 7, in press.

Gentner, D., & Stevens, A. (Eds.), *Mental models*. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1983.

Hayes-Roth, B., & Hayes-Roth, F. Concept learning and the recognition and classification of exemplars. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 321-338.

Hewitt, C. How to use what you know. *Proceedings of the Fourth International Joint Conference on Artificial Intelligence*, 1975, 189-198.

Hinton, G. E., & Anderson, J. A. (Eds.), *Parallel models of associative memory*. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1981.

Just, M. A., & Carpenter, P. A. Eye fixations and cognitive processes. *Cognitive Psychology*, 1976, *8*, 441-480.

Kay, A. C. Microelectronics and the personal computer. *Scientific American*, 1977, *237*, 231-244.

Kintsch, W. Notes on the structure of semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization and memory*. New York: Academic Press, 1972.

Kintsch, W. Comprehension and memory of text. In W. K. Estes, (Ed.), Handbook of learning and cognitive processes, vol. 6. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1978.

Kosslyn, S. M. Information representation in visual images. *Cognitive Psychology*, 1975, *7*, 341-370.

Kosslyn, S. M. *Image and mind*. Cambridge, Mass.: Harvard University Press, 1980.

Kosslyn, S. M., & Schwartz, S. P. Visual images as spatial representations in active memory. In E. M. Riseman & A. R. Hanson (Eds.), *Computer vision systems*.

Levin, J. A. *Network representation and rotation of letters*. Unpublished manuscript, University of California, San Diego, September, 1973. (Briefly discussed in Shepard & Cooper, 1982, pp. 141 and 144.)

Lindsey, P. H., & Norman, D. A. *Human information pjrocessing*. New York: Academic Press, 1972.

Maclean, J., & Schulman, G. The construction and maintenance of expectancies. *Quarterly Journal of Experimental Psychology*, 1978, *30*, 441-454.

McCarthy, J. Programs with common sense, in *Proceedings of the Teddington Conference on the Mechanisation of Thought Processes*. London: H. M. Stationery Office, 1960

McClelland, J. L. Retrieving general and specific information from stored knowledge of specifics. Proceedings of the Third Annual Conference of the Cognitive Science Society, 1981, 170-172.

McClelland, J. L., & Rumelhart, D. E. An interactive activation model of context effects in letter perception, Part 1: An account of basic findings. *Psychological Review*, 1981, *88*, 375-407.

Metzler, J. & Shepard, R. N. Transformational studies of the internal representation of three-dimensional objects. In R. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium*, Hillsdale, N. J.: Lawrence Erlbaum Associates, 1974.

Meyer, D. E. On the representation and retrieval of stored semantic information. *Cognitive Psychology*, 1970, *1*, 242-299.

Meyer, D. E., & Schvaneveldt, R. W. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 1971, *90*, 227-234.

Minsky, M. A framework for representing knowledge. In Winston, P. (Ed.), *The psychology of computer vision*. New York: McGraw-Hill, 1975.

Neely, J. H. Semantic priming and retrieval from lexical memory: Evidence for facilitory and inhibitory processes. *Memory & cognition*, 1976, *4*, 648-654.

Neely, J. H. Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, 1977, *106*, 226-254.

Newell, A. Production systems: Models of control structure. In W. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.

Newell, A. Physical symbol systems. In D. A. Norman (Ed.), *Perspectives on cognitive science*. Norwood, N. J.: Ablex Publishing Corp., 1981.

Newell, A., & Simon, H., GPS, a program that simulates human thought. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought*. New York: McGraw-Hill, 1963.

Newell, A., & Simon, H. *Human problem solving*. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Newport, E. L., & Bellugi, U. Linguistic expression of category levels in a visual-gestural language: A flower is a flower is a flower. In E. Rosch & B. B. Lloyd, (Eds.), *Cognition and categorization*. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1978.

Nilsson, N. J. *Principles of artificial intelligence*. Palo Alto, Calif.: Tioga, 1980.

Norman, D. A., & Bobrow, D. G. On the role of active memory processes in perception and cognition. In C. N. Cofer (Ed.), *The structure of human memory*. San Francisco: Freeman, 1976.

Norman, D. A., & Bobrow, D. G. Descriptions: An intermediate stage in memory retrieval. *Cognitive Psychology*, 1979, *11*, 107-123.

Norman, D. A., & Rumelhart, D. E. *Explorations in cognition*. San Francisco: Freeman, 1975.

Oden, G. C. Integration of fuzzy logical information. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 565-575.

Ortony, A. Beyond literal similarity. *Psychological Review*, 1979, *86*, 161 - 180.

Palmer, S. E. Fundamental aspects of cognitive representation. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Piaget, J. [*The origins of intelligence in children*] (M. Cook, trans.). New York: International Universities Press, 1952.

Posner, M. I., & Keele, S. W. On the genesis of abstract ideas. *Journal of Experimental Psychology*, 1968, *77*, 353-363.

Posner, M. I., & Snyder, C. R. R. Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition: The Loyola symposium*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.

Post, E. Formal reductions of the general combinatorial problem. *American Journal of Mathematics*, 1943, *65*, 197-268.

Quillian, M. R. Semantic memory. In M. Minsky (Ed.), *Semantic information processing*. Cambridge, Mass.: MIT Press, 1968.

Quillian, M. R. The teachable language comprehender. *Communications of the Association for Computing Machinery*, 1969, *12*, 459-475.

Reder, L. M. & Anderson, J. R. A partial resolution of the paradox of interference: The role of integrating knowledge. *Cognitive Psychology*, 1980, *12*, 447-472.

Reder, L. M., & Ross, B. H. Integrated knowledge in different tasks: The role of retrieval strategy on fan effects. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 1983, *9*, 55-72.

Rips, L. J., Shoben, E. J., & Smith, E. E. Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 1-20.

Rosch, E., Principles of categorization. In E. Rosch & B. B. Lloyd, (Eds.), *Cognition and categorization*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Rosch, E., & Lloyd, B. B. (Eds.), *Cognition and categorization*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Rosch, E., & Mervis, C. B., Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, 1975, *7*, 573-605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. Basic objects in natural categories. *Cogni-*

*tive Psychology*, 1976, *8*, 382-439.

Rumelhart, D. E., Notes on a schema for stories. In D. G. Bobrow & A. M. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.

Rumelhart, D. E. Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance VI*. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1977.

Rumelhart, D. E. *Understanding understanding* (Tech. Rep. CHIP 100). La Jolla, CA: University of California, San Diego, Center for Human Information Processing, January, 1981.

Rumelhart, D. E., & Levin, J. A. A language comprehension system. In D. A. Norman & D. E. Rumelhart, *Explorations in cognition*. San Francisco: Freeman, 1975.

Rumelhart, D. E., & McClelland, J. L. An interactive activation model of context effects in letter perception, Part 2: The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 1982 *89*, 60-94. 1982

Rumelhart, D. E., & Norman, D. A. Accretion, tuning, and restructuring: Three modes of learning. In J. W. Cotton & R. L. Klatzky (Eds.), *Semantic factors in cognition*. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1978.

Rumelhart, D. E., & Ortony, A. The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1977.

Schank, R. C., *Conceptual information processing*. New York: North-Holland, 1975.

Schank, R. C., Language and memory. *Cognitive Science*, 1980, *4*, 243-284.

Schank, R. C., Language and memory. In D. A. Norman (Ed.), *Perspectives on cognitive science*. Norwood, N. J.: Ablex, 1981. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1981.

Schank, R., & Abelson, R. *Scripts, Plans, Goals, and Understanding*. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1977.

Selfridge, O. Pandemonium: A paradigm for learning. In *Symposium on the mechanization of thought processes*. London: HM Stationery Office, 1959.

Shepard, R. N., & Cooper, L. A. *Mental images and their transformations*. Cambridge, Mass.: MIT Press, 1982

Shepard, R. N. & Feng, C. A chronometric study of mental paper folding. *Cognitive Psychology*, 1972, *3*, 228-243.

Smith, E. E., Adams, N., & Schorr, D. Fact retrieval and the paradox of interference. *Cognitive Psychology*, 1978, *10*, 438-464.

Smith, E. E., Balzano, G. J., & Walker, J. H. Nominal, perceptual, and semantic codes in picture categorization. In J. Cotton & R. Klatzky (Eds.), *Semantic factors in cognition*. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1978

Smith, E. E., & Medin, D. L., *Categories and concepts*. Cambridge, Mass.: Harvard University Press, 1981.

Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and process in semantic memory: A feature model for semantic decisions. *Psychological Review*, *81*, 214-241.

Tulving, E., & Donaldson, W. (Eds.), *Organization and memory*. New York: Academic Press, 1972.

Tversky, A. Features of similarity. *Psychological Review*, 1977, *84*, 327-352.

Tversky, A., & Gati, I. Studies of similarity. In E. Rosch & B. B. Lloyd, (Eds.), *Cognition and categorization*. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1978.

Waterman, D. A., & Hayes-Roth, F. (Eds.) *Pattern-directed inference systems*. New York: Academic Press, 1978.

Wilensky, R. *Understanding goal-based stories*. Ph.D. Thesis. Yale University, Computer Science Dept., Research Report 140, 1978.

Winograd T. *Understanding natural language*. New York: Academic Press, 1972.

Winograd, T. Frame representations and the declarative/procedural controversy. In D. G. Bobrow & A. M. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.

Wittgenstein, L. *Philosophical investigations*. Oxford: Blackwell, 1953.

Woods. W. A. What's in a link: Foundations for semantic networks. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. N. Y.: Academic Press, 1975.

Zadeh, L. Fuzzy sets. *Information and Control*, 1965, *8*, 338-353.

## CHIP Technical Report List

1. David M. Green and William J. McGill. On the equivalence of detection probabilities and well known statistical quantities. October, 1969.
2. Donald A. Norman. Comments on the information structure of memory. October, 1969.
3. Norman H. Anderson. Functional measurement and psychophysical judgment. October, 1969.
4. James C. Shanteau. An additive decision-making model for sequential estimation and inference judgments. October, 1969.
5. Norman H. Anderson. Averaging model applied to the size-weight illusion. October, 1969.
6. Norman H. Anderson and James C. Shanteau. Information integration in risky decision making. November, 1969.
7. George Mandler, Richard H. Meltzer, Zena Pearlstone. The structure of recognition. Effects of list tags and of acoustic and semantic confusion. November, 1969.
8. Dominic W. Massaro. Perceptual processes and forgetting in memory tasks. January, 1970.
9. Daniel Graboi. Searching for targets: The effects of specific practice. February, 1970.
10. James H. Patterson and David M. Green. Discrimination of transient signals having identical energy spectra. February, 1970.
11. Donald A. Norman. Remembrance of things past. June, 1970.
12. Norman H. Anderson. Integration theory and attitude change. August, 1970.
13. A.D. Baddeley and J.R. Ecob. Reaction time and short-term memory: A trace strength alternative to the high-speed exhaustive scanning hypothesis. November, 1970.
14. A.D. Baddeley. Retrieval rules and semantic coding in short-term memory. December, 1970.
15. Roy D. Patterson. Residue pitch as a function of the number and relative phase of the component sinusoids. March, 1971.
16. George Mandler and Marilyn A. Borges. Effects of list differentiation, category membership and prior recall on recognition. May, 1971.
17. David E. Rumelhart, Peter H. Lindsay, and Donald A. Norman. A process model for long-term memory. May, 1971.
18. David E. Rumelhart and Adele A. Abrahamson. Toward a theory of analogical reasoning. July, 1971.
19. Martin F. Kaplan. How response dispositions integrate with stimulus information. August, 1971.
20. Martin F. Kaplan and Norman H. Anderson. Comparison of information integration and reinforcement models for interpersonal attraction. August, 1971.
21. David M. Green and R. Duncan Luce. Speed-accuracy tradeoff in auditory detection. September, 1971.
22. David E. Rumelhart. A multicomponent theory of confusion among briefly exposed alphabetic characters. November, 1971.
23. Norman H. Anderson and Arthur J. Farkas. New light on order effects in attitude change. March, 1972.
24. Norman H. Anderson. Information integration theory: A brief survey. April, 1972.
25. Donald A. Norman. Memory, knowledge, and the answering of questions. May, 1972.
26. David J. Weiss. Averaging: An empirical validity criterion for magnitude estimation.
    Norman H. Anderson. Cross-task validation of functional measurement. June, 1972.
27. David E. Rumelhart and Patricia Siple. The process of recognizing tachistoscopically presented words. August, 1972.
28. Ebbe B. Ebbesen and Richard Bowers. The effects of proportion of risky to conservative arguments in a group discussion on risky shift. September, 1972.
29. Ebbe B. Ebbesen and Michael Haney. Flirting with death: Variables affecting risk taking on our nation's highways. September, 1972.
30. Norman H. Anderson. Algebraic models in perception. November, 1972.
31. Norman H. Anderson. Cognitive algebra: Information integration applied to social attribution. December, 1972.
32. Jean M. Mandler and Nancy L. Stein. Recall recognition of pictures by children as a function of organization and of distractor similarity. January, 1973.
33. David E. Rumelhart and Donald A. Norman. Active semantic networks as a model of human memory.
    Marc Eisenstadt and Yaakov Kareev. Towards a model of human game playing. June, 1973.
34. George Mandler. Memory storage and retrieval: Some limits on the reach of attention and consciousness. July, 1973.
35. Kent L. Norman. A method of maximum likelihood estimation for stimulus integration theory. August, 1973.
36. Yaakov Kareev. A model of human game playing. August, 1973.
37. Donald A. Norman. Cognitive organization and learning. August, 1973.
38. The Center for Human Information Processing: A Five Year Report — 1968-73.

39. Larry D. Rosen and J. Edward Russo. Binary processing in multi-alternative choice. October, 1973.

40. Samuel Himmelfarb and Norman H. Anderson. Integration theory analysis of opinion attribution. December, 1973.

41. George Mandler. Consciousness: Respectable, useful, and probably necessary. March, 1974.

42. Norman H. Anderson. The problem of change-of-meaning. June, 1974.

43. Norman H. Anderson. Methods for studying information integration. June, 1974.

44. Norman H. Anderson. Basic experiments in person perception. June, 1974.

45. Norman H. Anderson. Algebraic models for information integration. June, 1974.

46. Ebbe B. Ebbesen and Vladimir J. Konečni. Cognitive algebra in legal decision making. September, 1974.

47. Norman H. Anderson. Equity judgments as information integration.
    Arthur J. Farkas and Norman H. Anderson. Input summation and equity summation in multi-cue equity judgments. December, 1974.

48. George Mandler and Arthur Graesser II. Dimensional analysis and the locus of organization. January, 1975.

49. James L. McClelland. Preliminary letter identification in the perception of words and nonwords. April, 1975.

50. Donald A. Norman and Daniel G. Bobrow. On the role of active memory processes in perception and cognition. May, 1975.

51. J. Edward Russo. The value of unit price information. An information processing analysis of point-of-purchase decisions. June, 1975.

52. Elissa L. Newport. Motherese: The speech of mothers to young children. August, 1975.

53. Norman H. Anderson and Cheryl C. Graesser. An information integration analysis of attitude change in group discussion. September, 1975.

54. Lynn A. Cooper. Demonstration of a mental analog of an external rotation.
    Lynn A. Cooper and Peter Podgorny. Mental transformations and visual comparison processes: Effects of complexity and similarity. October, 1975.

55. David E. Rumelhart and Andrew Ortony. The representation of knowledge in memory. January, 1976.

56. David E. Rumelhart. Toward an interactive model of reading. March, 1976.

57. Jean M. Mandler, Nancy S. Johnson, and Marsha DeForest. A structural analysis of stories and their recall: From "Once upon a time" to "Happily ever after". March, 1976.

58. David E. Rumelhart. Understanding and summarizing brief stories. April, 1976.

59. Lynn A. Cooper and Roger N. Shepard. Transformations on representations of objects in space. April, 1976.

60. George Mandler. Some attempts to study the rotation and reversal of integrated motor patterns. May, 1976.

61. Norman H. Anderson. Problems in using analysis of variance in balance theory. June, 1976.

62. Norman H. Anderson. Social perception and cognition. July, 1976.

63. David E. Rumelhart and Donald A. Norman. Accretion, tuning and restructuring: Three modes of learning. August, 1976.

64. George Mandler. Memory research reconsidered: A critical view of traditional methods and distinctions. September, 1976.

65. Norman H. Anderson and Michael D. Klitzner. Measurement of motivation.
    Michael D. Klitzner and Norman H. Anderson. Motivation x expectancy x value: A functional measurement approach. November, 1976.

66. Vladimir J. Konečni. Some social, emotional, and cognitive determinants of aesthetic preference for melodies in complexity. December, 1976.

67. Hugh Mehan, Courtney B. Cazden, LaDonna Coles, Sue Fisher, Nick Maroules. The social organization of classroom lessons. December, 1976.

67a. Hugh Mehan, Courtney B. Cazden, LaDonna Coles, Sue Fisher, Nick Maroules. Appendices to the social organization of classroom lessons. December, 1976.

68. Norman H. Anderson. Integration theory applied to cognitive responses and attitudes. December, 1976.

69. Norman H. Anderson and Diane O. Cuneo. The height + width rule in children's judgments of quantity. June, 1977.
    Norman H. Anderson and Clifford H. Butzin. Children's judgments of equity. June, 1977.

70. Donald R. Gentner and Donald A. Norman. The FLOW tutor: Schemas for tutoring. June, 1977.

71. George Mandler. Organization and repetition: An extension of organizational principles with special reference to rote learning. May, 1977.

72. Manuel Leon. Coordination of intent and consequence information in children's moral judgements. August, 1977.

73. Ted Supalla and Elissa L. Newport. How many seats in a chair? The derivation of nouns and verbs in American Sign Language. November, 1977.

74. Donald A. Norman and Daniel G. Bobrow. Descriptions: A basis for memory acquisition and retrieval. November, 1977.

75. Michael D. Williams. The process of retrieval from very long term memory. September, 1978.

76. Jean M. Mandler. Categorical and schematic organization in memory. October, 1978.

77. James L. McClelland. On time relations of mental processes: A framework for analyzing processes in cascade. October, 1978.

78. Jean M. Mandler and Marsha DeForest. Developmental invariance in story recall. November, 1978.

79. David E. Rumelhart. Schemata: The building blocks of cognition. December, 1978.

80. Nancy S. Johnson and Jean M. Mandler. A tale of two structures: Underlying and surface forms in stories. January, 1979.

81. David E. Rumelhart. Analogical processes and procedural representations. February, 1979.

82. Ross A. Bott. A study of complex learning: Theory and methodologies. March, 1979.

83. Laboratory of Comparative Human Cognition. Toward a unified approach to problems of culture and cognition. May, 1979.

84. George Mandler and Lawrence W. Barsalou. Steady state memory: What does the one-shot experiment assess? May, 1979.

85. Norman H. Anderson. Introduction to cognitive algebra. June, 1979.

86. Edited by Michael Cole, Edwin Hutchins, James Levin and Naomi Miyake. Naturalistic problem solving and microcomputers. Report of a Conference. June, 1979.

87. Donald A. Norman. Twelve issues for cognitive science. October, 1979.

88. Donald A. Norman. Slips of the mind and an outline for a theory of action. November, 1979.

89. The Center for Human Information Processing: A Description and a Five-Year Report (1974-1979). November, 1979.

90. Michael Cole and Peg Griffin. Cultural amplifiers reconsidered. December, 1979.

91. James L. McClelland and David E. Rumelhart. An interactive activation model of the effect of context in perception. Part I. April, 1980.

92. James L. McClelland and J.K. O'Regan. The role of expectations in the use of peripheral visual information in reading. February, 1980.

93. Edwin Hutchins. Conceptual structures of Caroline Island navigation. May, 1980.

94. Friedrich Wilkening and Norman H. Anderson. Comparison of two rule assessment methodologies for studying cognitive development. June, 1980.

95. David E. Rumelhart and James L. McClelland. An interactive activation model of the effect of context in perception. Part II. August, 1980.

96. Jean M. Mandler. Structural invariants in development. September, 1980.

97. David E. Rumelhart and Donald A. Norman. Analogical processes in learning. October, 1980.

98. James A. Levin and Yaakov Kareev. Personal computers and education: The challenge to schools. November, 1980.

99. Donald A. Norman and Tim Shallice. Attention to action: Willed and automatic control of behavior. December, 1980.

100. David E. Rumelhart. Understanding understanding. January, 1981.

101. George Mandler. The structure of value: Accounting for taste. May, 1981.

102. David E. Rumelhart and Donald A. Norman. Simulating a skilled typist: A study of skilled cognitive-motor performance. May, 1981.

103. Jean M. Mandler. Representation. June, 1981.

104. Donald R. Gentner. Skilled finger movements in typing. July, 1981.

105. Edwin L. Hutchins and James A. Levin. Point of view in problem solving. August, 1981.

106. Michael Cole. Society, mind and development. September, 1981.
Michael Cole. The zone of proximal development: Where culture and cognition create each other. September, 1981.

107. Laboratory of Comparative Human Cognition. Culture and cognitive development. November, 1981.

108. Donald R. Gentner. Evidence against a central control model of timing in typing. December, 1981.

109. Robert M. Boynton and Allen L. Nagy. The La Jolla analytic colorimeter: Optics, calibrations, procedures, and control experiments. December, 1981.

110. Jonathan T. Grudin and Serge Larochelle. Digraph frequency effects in skilled typing. February, 1982.

111. Donald R. Gentner, Jonathan T. Grudin, Serge Larochelle, Donald A. Norman, David E. Rumelhart. Studies of typing from the LNR Research Group. September, 1982.

112. Donald A. Norman. Five papers on human-machine interaction. May, 1982.

113. Naomi Miyake. Constructive interaction. June, 1982.

114. Donald Gentner. The development of typewriting skill. September, 1982.

115. Margaret M. Riel. Investigating the system of development: The skills and abilities of dysphasic children. February, 1983.

116. David E. Rumelhart and Donald A. Norman. Representation in memory. June, 1983.

Note: Requests for CHIP reports should be addressed to the author. Reports are also available through the Library Loan Service of the University of California, San Diego, La Jolla, California 92093.